



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 100034

TO: Anish Gupta
Location: cm-1/9A13/11D13
Art Unit: 1654

July 30, 2003

Case Serial Number: 09/787070

From: P. Sheppard
Location: CM1-1E03
Phone: (703) 308-4499

sheppard@uspto.gov

Search Notes

THIS PAGE BLANK (USPTO)

STIC-Bi tech/Ch mLib

100034

From: Gupta, Anish
Sent: Tuesday, July 29, 2003 3:53 PM
To: STIC-Biotech/ChemLib
Subject: search request

Application Serial Number: 09/787,070
location: 9A13 Art Unit: 1654

Date: 7-27-03
phone: 308-4001

Please search seq. ID NO. 1-8

thank you
anish gupta

Point of Contact
P. Sheppard
Telephone number: (703) 308-4499

7/30/03

Point of Contact
P. Sheppard
Telephone number: (703) 308-4499

RECEIVED
JUL 29 2003
STIC

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 27.1429 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-1

Perfect score: 92
Sequence: 1 VYQHQAAMKPMIQPKT 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

```

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP_human: *
5: SP_invertebrate: *
6: SP_mammal: *
7: SP_mhc: *
8: SP_organelle: *
9: SP_phage: *
10: SP_plant: *
11: SP_rodent: *
12: SP_virus: *
13: SP_vertebrate: *
14: SP_unclassified: *
15: SP_virus: *
16: SP_bacteriap: *
17: SP_archaeap: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	81	88.0	210	6	062825	062825 bubalus bub
2	77	83.0	223	6	09MYU6	09MYU6 capra hircu
3	68	73.9	223	6	09TQ07	09TQ07 capra hircu
4	68	73.9	223	6	09GK07	09GK07 capra hircu
5	50	54.3	693	11	09WV66	09WV66 mus muscul
6	48	52.2	225	4	09EMR9	09EMR9 homo sapien
7	48	52.2	328	16	08CP60	08CP60 staphylococ
8	48	52.2	418	16	09MYW5	09MYW5 thermotoga
9	48	52.2	704	4	09H9J2	09H9J2 homo sapien
10	46.5	50.5	353	5	001557	001557 caenorhabdi
11	46.5	50.5	354	5	001555	001555 caenorhabdi
12	46	50.0	265	8	09ZRR2	09ZRR2 cyanidorsch
13	46	50.0	458	4	08WVP9	08WVP9 homo sapien
14	46	50.0	475	4	09FKK5	09FKK5 homo sapien
15	45	50.0	893	4	09C0B8	09C0B8 homo sapien
16	45	48.9	309	10	09FCU9	09FCU9 arabidopsis

17	45	48.9	311	5	Q19257	Q19257 caenorrhadi
18	45	48.9	931	5	Q95H48	Q95H48 cyza sattiv
19	44.5	48.4	205	13	Q9PM95	Q9PM95 anas platyr
20	44	47.8	85	10	Q81534	Q81534 saccharin o
21	44	47.8	80	10	Q81399	Q81399 orobanch e
22	44	47.8	328	17	Q9UYA2	Q9UYA2 pyrococcus
23	44	47.8	340	10	Q96568	Q96568 ipomoea puru
24	44	47.8	379	10	Q04065	Q04065 perilla fru
25	44	47.8	386	10	Q23923	Q23923 digitalis l
26	44	47.8	388	10	Q9FS37	Q9FS37 coretia hyb
27	44	47.8	389	10	Q8LE19	Q8LE19 niterembegi
28	44	47.8	389	10	Q43040	Q43040 petunia hyb
29	44	47.8	389	10	Q93XP8	Q93XP8 nicotiana t
30	44	47.8	389	10	Q9MSB2	Q9MSB2 petunia hyb
31	44	47.8	390	10	Q9LRB2	Q9LRB2 scutellaria
32	44	47.8	390	10	Q48564	Q48564 scutellaria
33	44	47.8	390	10	Q02122	Q02122 scutellaria
34	43	46.7	69	2	P74958	P74958 shewanella
35	43	46.7	136	5	Q25047	Q25047 haemochus
36	43	46.7	136	5	Q25044	Q25044 haemochus
37	43	46.7	161	5	Q26266	Q26266 caenorrhadi
38	43	46.7	180	5	Q25045	Q25045 haemochus
39	43	46.7	205	5	Q9N6M8	Q9N6M8 dirosophila
40	43	46.7	279	8	Q9G3M0	Q9G3M0 arabidopsis
41	43	46.7	279	10	Q8LBB2	Q8LBB2 arabidopsis
42	43	46.7	279	10	Q8LBB2	Q8LBB2 arabidopsis
43	43	46.7	297	5	Q44075	Q44075 caenorrhadi
44	43	46.7	308	2	Q9REJ7	Q9REJ7 agrobacteri
45	43	46.7	393	11	Q91XA2	Q91XA2 mus musculu

ALIGNMENTS

RESULT 1

ID	062825	PRELIMINARY;	PRT;	210	AA
----	--------	--------------	------	-----	----

AC	062825	
DT	01-AUG-1998	(TREMBLrel. 07, Created)
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)
DE	A52-casein (Fragment) .	
OS	Bubalus bubalis (Domestic water buffalo) .	
OC	Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora;	
OC	Bovidae; Bovinae; Bubalus.	
OX	NCBI_TaxId=89462;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Mammary gland;	
RA	Das P., Jain S., Garg L.C. ;	
RT	"Cloning and nucleotide sequence of cDNA encoding A52-casein in B	
RL	bubalis." ;	
RD	Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AJ005431; CAA06534.2 ; -	
DR	InterPro; IPR001588; Casein.	
DR	Pfam; PF000363; caseins; 2.	
FT	NON_TER	
QO	SEQUENCE	210 AA; 24700 MW; 05DEF95963F1132C CRC64;

Query Match	88.0%	Score	81	DB	6	length	210
Best Local Similarity	87.5%	Pred.	NO.	1	Se	05	
Matches	14	Conservative	1	Mismatches	1	Indels	0
						Gaps	0

```
Qy      1 VYQHOKAMKPWTQPRT 16
        |||:||||||| ||||
Db      186 VYQYQKAMKPWTQPKT 201
```

RESULT 2	
Q9MYU6	
ID Q9MYU6	PRELIMINARY;
AC Q9MYU6;	PRT; 223 AA

AC Q9MYU6;

DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Alpha s2-casein.
 OS CSM152.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OC NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lagomigro R., Pilla F., Mataasino D., Zullo A.;
 RT "A new allele of goat alpha s2-casein gene."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ249716; CAB94236.1; -
 DR InterPro: IPR001588; Casein.
 DR Pfam: PF00363; caseins; 2.
 DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
 SQ SEQUENCE 223 AA; 26403 MW; 0E1FE83F24DA85E2 CRC64;
 Query Match 83.7%; Score 77; DB 6; Length 223;
 Best Local Similarity 87.5%; Pred. No. 7.5e-05;
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 VYOHOKAMKRWIOPT 16
 Db 199 VDOHOKAMKRWIOPT 214
 RESULT 3
 Q9TT07 PRELIMINARY; PRT; 223 AA.
 ID Q9TT07
 AC Q9TT07
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Alpha s2-casein.
 GN CSM152.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OC NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veltre C.C., Pilla F.F., Lagomigro R.R.;
 RT "A new allele of goat alpha s2-casein."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ249995; CAB59920.1; -
 DR InterPro: IPR001588; Casein.
 DR Pfam: PF00363; caseins; 2.
 DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
 SQ SEQUENCE 223 AA; 26433 MW; CE9F4DC8D7688293 CRC64;
 Query Match 73.9%; Score 68; DB 6; Length 223;
 Best Local Similarity 81.2%; Pred. No. 0.0023;
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 VYOHOKAMKRWIOPT 16
 Db 199 VDOHOKAMKRWIOPT 214
 RESULT 4
 Q9GK07 PRELIMINARY; PRT; 223 AA.
 ID Q9GK07
 AC Q9GK07
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Alpha s2-casein.
 GN CSM152.

OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OC NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veltre C., Pilla F., Lagomigro R.;
 RT "A new allele of alpha s2-casein."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ297311; CAC21704.2; JOINED.
 DR EMBL: AJ297311; CAC21704.2; JOINED.
 DR EMBL: AJ242728; CAC21704.2; JOINED.
 DR EMBL: AJ297312; CAC21704.2; JOINED.
 DR EMBL: AJ297313; CAC21704.2; JOINED.
 DR EMBL: AJ297314; CAC21704.2; JOINED.
 DR EMBL: AJ242527; CAC21704.2; JOINED.
 DR EMBL: AJ297315; CAC21704.2; JOINED.
 DR EMBL: AJ297316; CAC21704.2; JOINED.
 DR EMBL: AJ242526; CAC21704.2; JOINED.
 DR EMBL: AJ242528; CAC21704.2; JOINED.
 DR EMBL: AJ242533; CAC21704.2; JOINED.
 DR InterPro: IPR001588; Casein.
 DR Pfam: PF00363; caseins; 2.
 DR PROSITE: PS00306; CASEIN_ALPHA_BETA; 1.
 SQ SEQUENCE 223 AA; 26433 MW; CE9765E8D7688C9D CRC64;
 Query Match 73.9%; Score 68; DB 6; Length 223;
 Best Local Similarity 81.2%; Pred. No. 0.0023;
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 VYOHOKAMKRWIOPT 16
 Db 199 VDOHOKAMKRWIOPT 214
 RESULT 5
 Q9WV66 PRELIMINARY; PRT; 693 AA.
 ID Q9WV66
 AC Q9WV66
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Axotrophin.
 GN AXOT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Haendel M.A., Wagner M.K., Lyons G.E.;
 RT "Insertional mutagenesis of a novel gene, axotrophin, causes callosal
 RT agenesis and neural degeneration."
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF155739; AAD38411.1; -
 DR EMBL: BC025029; AAH25029.1; -
 DR MGD; MGI:1931053; Axot.
 SQ SEQUENCE 693 AA; 76598 MW; 7082766DE39E4CED CRC64;
 Query Match 54.3%; Score 50; DB 11; Length 693;
 Best Local Similarity 57.1%; Pred. No. 7;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 Qy 2 YOHOKAMKRWIOPT 15
 Db 580 YVHOCMKRWIOPT 593

```

RESULT 6
Q9BTR9 PRELIMINARY; PRT; 225 AA.
AC Q9BTR9;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003404; AAH03404.1; -.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 225 AA; 25142 MW; 261C78A4FFFA66E CRC64;

Query Match
Best Local Similarity 52.2%; Score 48; DB 4; Length 225;
Pred. No. 4.8;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 YOHOKAKMPWIOPK 15
Db 99 YVHODCKMKWLOAK 112

RESULT 7
Q8CP60 PRELIMINARY; PRT; 328 AA.
AC Q8CP60;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DE Thioedoxin reductase-like protein.
GN SELL68.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016748; AAC04767.1; -.
KW Complete proteome.
SQ SEQUENCE 328 AA; 36814 MW; 286F6B34EBF6F87 CRC64;

Query Match
Best Local Similarity 52.2%; Score 48; DB 16; Length 328;
Pred. No. 7;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 QHOKAKMPWIOPK 14
Db 190 QYFKAIPWLP 201

RESULT 8
Q9WYWS PRELIMINARY; PRT; 418 AA.
AC Q9WYWS;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DE PHO-related protein.
GN TM0495.
OS Thermotoga maritima.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.

```

```

OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.; and Bacteria from
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima.";
DR EMBL; AE001726; AAD35580.1; -.
DR TIGR; TM0495; -.
DR InterPro; IPR003714; PhOH.
DR InterPro; IPR006596; PINC.
DR Pfam; PF02562; PhOH; 1.
DR SMART; SM00670; PINC; 1.
KW Complete proteome.
SQ SEQUENCE 418 AA; 47518 MW; A42EB74F86D7C8FB CRC64;

Query Match
Best Local Similarity 52.2%; Score 48; DB 16; Length 418;
Pred. No. 9;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 QKAKMPWIOPK 14
Db 285 EKMKRKPWLP 294

RESULT 9
Q9H992 PRELIMINARY; PRT; 704 AA.
AC Q9H992;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE Hypothetical protein FLJ12911.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuno Y., Sasaki N.;
RL "NEO human cDNA sequencing project.";
DR EMBL; AK022973; BAB14340.1; -.
DR Gene; HGNC:11793; AXOT.
KW Hypothetical protein.
SQ SEQUENCE 704 AA; 78050 MW; 1271F74D7F389130 CRC64;

Query Match
Best Local Similarity 52.2%; Score 48; DB 4; Length 704;
Pred. No. 15;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 YOHOKAKMPWIOPK 15
Db 578 YVHODCKMKWLOAK 591

RESULT 10
Q01557 PRELIMINARY; PRT; 363 AA.
AC Q01557;
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)

```

DE F47B3.1 Protein.
 GN F47B3.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Kirsten J., Laister N., Latreille P.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lighting J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Thierly-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z., Le T.T.;
 RT "The sequence of C. elegans cosmid F47B3.";
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DR HSSP; P29350; 1GMZ.
 DR WormPep; F47B3.1; CE10616.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; TYR_PP.
 DR Pfam; PF00102; Y_PRTYPHPTASE.
 DR PRINTS; PR00700; PRTYPHPTASE.
 DR SMART; SM00194; PTPc; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 KM Hydrolase.
 SQ SEQUENCE 363 AA; 40796 MW; 58AF84204CBDCD011 CRC64;
 Query Match 50.5%; Score 46.5; DB 5; Length 363;
 Best Local Similarity 69.2%; Pred. No. 14;
 Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
 QY 4 HOKAM---KPMIQ 13
 DB 48 HOKAMIVKPMVQ 60

RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Kirsten J., Laister N., Latreille P.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lighting J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Thierly-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z., Le T.T.;
 RT "The sequence of C. elegans cosmid F47B3.";
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 DR HSSP; P29350; 1GMZ.
 DR WormPep; F47B3.1; CE10626.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; TYR_PP.
 DR Pfam; PF00102; Y_PRTYPHPTASE.
 DR PRINTS; PR00700; PRTYPHPTASE.
 DR SMART; SM00194; PTPc; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 KM Hydrolase.
 SQ SEQUENCE 364 AA; 40860 MW; B2B79C1D43F44B8 CRC64;
 Query Match 50.5%; Score 46.5; DB 5; Length 364;
 Best Local Similarity 69.2%; Pred. No. 14;
 Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;
 QY 4 HOKAM---KPMIQ 13
 DB 49 HOKAMIVKPMVQ 61

RESULT 12
 Q9ZZR2
 ID Q9ZZR2 PRELIMINARY; PRT; 265 AA.
 AC Q9ZZR2;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Succinate dehydrogenase iron-sulfur protein.
 GN SDHB.
 OS Cyathioschyron merolae.
 OC Mitochondrion.
 OC Eukaryota; Rhodophyta; Bangiophyceae; Cyathioschyron.
 OX NCBI_TaxID=45157;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=10D;
 RX MEDLINE=99030526; PubMed=9801318;
 RA Ohta N., Sato N., Kuroiwa T.;
 RT "Structure and organization of the mitochondrial genome of the
 RT unicellular red alga Cyathioschyron merolae deduced from the complete
 RT nucleotide sequence.";
 RL Nucleic Acids Res. 26:5190-5198 (1998).
 CC -1- COFACTOR: BINDS 1 ZFE-25 CLUSTER (BY SIMILARITY).
 DR EMBL; D89861; BAA34653.1; -.
 DR HSSP; P00364; 1KF6.

DR InterPro: IPR006058; 2Fe2S-ferredoxin.
 DR InterPro: IPR001450; 4Fe4S-ferredoxin.
 DR InterPro: IPR004489; DhsB.
 DR InterPro: IPR001041; Ferredoxin.
 DR Pfam: PRO011; fer2; 1.
 DR TIGRfam: TIGR00384; dhsB; 1.
 DR PROSITE: PS00197; 2FE2S_FERREDOXIN; 1.
 DR PROSITE: PS00198; 4FE4S_FERREDOXIN; 1.
 DR Iron_sulfur; Mitochondrion.
 KW SEQUENCE 265 AA; 30924 MW; D3E2D87DC55497D9 CRC64;

Query Match 50.0%; Score 46; DB 8; Length 265;
 Best Local Similarity 46.2%; Pred. No. 12;
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 YOHOKAMKWP10 14
 DB 134 YTOHKSIOFWLP 146

RESULT 13

O8WVP9

ID O8WVP9; PRELIMINARY; PRT; 458 AA.

AC O8WVP9; 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Hypothetical protein (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC017658; AA017658.1; -
 DR InterPro: IPR003659; Plexin-like.
 DR InterPro: IPR002165; Plexin_repeat.
 DR InterPro: IPR001627; Sema.
 DR Pfam; PF01437; PSI; 1.
 DR Pfam; PF01403; Sema; 1.
 DR SMART; SM00423; PSI; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 458 AA; 50263 MW; B8A52233DC848E7 CRC64;

Query Match 50.0%; Score 46; DB 4; Length 458;
 Best Local Similarity 61.5%; Pred. No. 21;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YOHOKAMKWP10 13
 DB 176 LYOPOLATRPW10 188

RESULT 14

O96FK5

ID O96FK5; PRELIMINARY; PRT; 475 AA.

DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RA Strausberg R.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC017071; AA017071.1; -
 DR InterPro: IPR003659; Plexin-like.
 DR InterPro: IPR002165; Plexin_repeat.
 DR InterPro: IPR001627; Sema.
 DR Pfam; PF01437; PSI; 1.
 DR Pfam; PF01403; Sema; 1.
 DR SMART; SM00423; PSI; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 475 AA; 52298 MW; EE90909701AE2264 CRC64;

Query Match 50.0%; Score 46; DB 4; Length 475;
 Best Local Similarity 61.5%; Pred. No. 22;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YOHOKAMKWP10 13
 DB 193 LYOPOLATRPW10 205

RESULT 15

O9COB8

ID O9COB8; PRELIMINARY; PRT; 893 AA.

AC O9COB8; 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Hypothetical protein KIAA1745 (Fragment).
 GN KIAA1745.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21082932; PubMed=11214970;
 RX Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Ref. 7:347-355 (2000).
 DR EMBL; AB051532; BAB21836.1; -
 DR InterPro: IPR003659; Plexin-like.
 DR InterPro: IPR002165; Plexin_repeat.
 DR InterPro: IPR001627; Sema.
 DR Pfam; PF01437; PSI; 1.
 DR Pfam; PF01403; Sema; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00630; Sema; 1.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 893 AA; 98361 MW; D6C6C48DE524F14 CRC64;

Query Match 50.0%; Score 46; DB 4; Length 893;
 Best Local Similarity 61.5%; Pred. No. 42;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YOHOKAMKWP10 13
 DB 611 LYOPOLATRPW10 623

Search completed: July 30, 2003, 16:29:20
 Job time : 29.1429 secs

THIS PAGE BLANK (uspto)


```

CC 50-59, OR 51-60.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -1- DATABASE: NAME=Protein Spotlight;
CC NOTE=Issue 16 of November 2001;
CC WWW="http://www.expasy.org/spotlight/articles/spot16.htm".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
CC -----
DR EMIL, M16644; AAA30479.1; -.
DR PIR, JQ2008; KABS02.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
KW Milk; Phosphorylation; Signal; Repeat; Antibiotic.
FT SIGNAL 1 15
FT CHAIN 16 222 ALPHA-S2 CASEIN.
FT PEPTIDE 165 203 CASOCIDIN-1.
FT REPEAT 76 140
FT REPEAT 158 222
FT MOD_RES 23 23 PHOSPHORYLATION.
FT MOD_RES 24 24 PHOSPHORYLATION.
FT MOD_RES 25 25 PHOSPHORYLATION.
FT MOD_RES 31 31 PHOSPHORYLATION.
FT MOD_RES 71 71 PHOSPHORYLATION.
FT MOD_RES 72 72 PHOSPHORYLATION.
FT MOD_RES 73 73 PHOSPHORYLATION.
FT MOD_RES 76 76 PHOSPHORYLATION.
FT MOD_RES 144 144 PHOSPHORYLATION.
FT MOD_RES 146 146 PHOSPHORYLATION.
FT CONFLICT 102 102 Q -> E (IN REF. 2).
FT FT SEQUENCE 222 AA; 26019 MW; 81E7408AF1C12F7C CRC64;
SQ
Query Match 100.0%; Score 92; DB 1; Length 222;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 VYOHOKAMKPTQPKT 16
Db 198 VYOHOKAMKPTQPKT 213
.
RESULT 2
CAS2_CAPHI
ID_CAS2_CAPHI STANDARD; PRT; 223 AA.
AC P13049;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-S2 casein precursor (Alpha-S2-CN).
GN CSNIS2.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Caprinae; Capra.
CX NCBI_TaxID:9925;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93216130; PubMed=8462880;
RA Bouniol C.;
RT "Sequence of the goat alpha s2-casein-encoding cDNA.";
RL Gene 125:235-236(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX Bouniol C., Brignon G., Mahe M.-F., Prinz C.;
RT "Characterization of goat allelic alpha-s2-caseins A and B: further
evidence of the phosphorylation code of caseins.";
RL Protein Seq. Data Anal. 5:213-218(1993).

```

```

RN [3] SEQUENCE FROM N.A. (VARIANT C).
RP MEDLINE=95030556; PubMed=7943951;
RX Bouniol C., Bignon G., Mahe M.F., Printz C.,
RT "Biochemical and genetic analysis of variant C of caprine alpha s2-
RT casein (Capra hircus)";
RL Anim. Genet. 25:173-177(1994).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- POLYMORPHISM: THREE ALLELES OF ALPHA-S2 CASEIN ARE KNOWN: A, B AND
CC C. THE FREQUENCIES OF THE ALLELES IS ESTIMATED TO BE 0.85, 0.04
CC AND 0.11 IN THE FRENCH DAIRY BREEDS 'ALPINE' AND 'SAANEN'.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X65160; CAA46278.1; -
DR EMBL; S74171; AAB32166.1; -
DR PIR; I46995; I46995.
DR PIR; S33881; JN0547.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
KM Milk; Phosphorylation; Signal; Repeat; Polymorphism.
KT SIGNAL 1 15
KT CHAIN 16 223
KT REPEAT 77 141
KT REPEAT 159 223
FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 77 77 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 145 145 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 147 147 PHOSPHORYLATION (BY SIMILARITY).
FT VARIANT 79 79 E -> K (IN VARIANT C).
FT VARIANT 182 182 K -> I (IN VARIANT C).
SQ SEQUENCE 223 AA; 26389 MW; 187DBF42FD688291 CRC64;

Query Match 83.7%; Score 77; DB 1; Length 223;
Best Local Similarity 87.5%; Pred. No. 2.6e-05;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYOHOKAMKWPLOPKT 16
| | | | | | | | | |
DB 199 VDOHOKAMKWPLOPKT 214

RESULT 3
CAS2_SHEEP STANDARD; PRT; 223 AA.
AC P04654;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-s2 casein precursor.
GN CSN1S2.
OS Ovis aries (sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OK NCBI_TaxID=9940;
RN [1]

```


RP SEQUENCE FROM N.A.
 RX MEDLINE=9610467; PubMed=3002499;
 RA Boissard M., Petrisant G.;
 RT "Complete sequence of ovine alpha s2-casein messenger RNA.";
 RL Biochimie 67:1043-1051(1985).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC CALCIUM PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X03238; CAA26983.1; -.
 CC PIR; A25070; KASHS2.
 CC InterPro; IPR001588; Casein.
 CC Pfam; PF00363; caseins; 2.
 CC PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 CC MILK; Phosphorylation; Repeat; Signal.
 CC KW
 CC FT SIGNAL 1 15
 CC FT CHAIN 16 223 ALPHA-S2 CASEIN.
 CC FT REPEAT 77 141
 CC FT REPEAT 159 223
 CC FT MOD_RES 23 23 PHOSPHORYLATION (POTENTIAL).
 CC FT MOD_RES 24 24 PHOSPHORYLATION (POTENTIAL).
 CC FT MOD_RES 25 25 PHOSPHORYLATION (POTENTIAL).
 CC FT MOD_RES 72 72 PHOSPHORYLATION (POTENTIAL).
 CC FT MOD_RES 73 73 PHOSPHORYLATION (POTENTIAL).
 CC FT MOD_RES 74 74 PHOSPHORYLATION (POTENTIAL).
 CC FT VARIANT 64 64 D -> N.
 CC FT SEQUENCE 223 AA; 26332 MW; 67212935E27426D7 CRC64;
 SQ
 Query Match 83.7%; Score 77; DB 1; Length 223;
 Best Local Similarity 87.5%; Pred. No. 2.6e-05;
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VYOHQKMKPWTQPKT 16
 Db 199 VDHQKMKPWTQPKT 214
 RESULT 4
 CAS2_PIG STANDARD; PRT; 235 AA.
 ID CAS2_PIG
 AC P39036;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Alpha-S2 casein precursor.
 GN CSN152.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sub.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=92367960; PubMed=1503276;
 RA Alexander L.J., Das Gupta N.A., Beattie C.W.;
 RT "The sequence of porcine alpha s2-casein cDNA.";
 RL Ant. Genet. 23:365-367(1992).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC CALCIUM PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X54975; CAA38719.1; -.
 CC InterPro; IPR001588; Casein.
 CC Pfam; PF00363; caseins; 2.
 CC PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 CC MILK; Phosphorylation; Signal.
 CC KW
 CC FT SIGNAL 1 15 BY SIMILARITY.
 CC FT CHAIN 16 235 ALPHA-S2 CASEIN.
 CC FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 70 70 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 71 71 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
 CC FT SEQUENCE 235 AA; 27570 MW; C903B760D184C14C CRC64;
 SQ
 Query Match 60.9%; Score 56; DB 1; Length 235;
 Best Local Similarity 68.8%; Pred. No. 0.064;
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VYOHQKMKPWTQPKT 16
 Db 208 VDHQKMKPWTQPKT 223
 RESULT 5
 CAS2_CAMDR STANDARD; PRT; 193 AA.
 ID CAS2_CAMDR
 AC O97944;
 DT 15-UTL-1999 (Rel. 38, Created)
 DT 15-UTL-1999 (Rel. 38, Last sequence update)
 DT 15-UTL-1999 (Rel. 38, Last annotation update)
 DE Alpha-S2 casein precursor.
 GN CSN152.
 OS Camelus dromedarius (Dromedary) (Arabian camel).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 OX NCBI_TaxID=9838;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=Somali; TISSUE=udder;
 RX MEDLINE=98291310; PubMed=9627840;
 RA Kapeller S., Farah Z., Puhon Z.;
 RT "Sequence analysis of Camelus dromedarius milk caseins.";
 RL J. Dairy Res. 65:209-222(1998).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC CALCIUM PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ016629; CAA10078.1; -.
 CC InterPro; IPR001588; Casein.
 CC Pfam; PF00363; caseins; 2.
 CC PROSITE; PS00306; CASEIN_ALPHA_BETA; FALSE_NEG.
 CC -----

DR EMBL; BC010701; AAH10701.1; ALT_INIT.
 DR EMBL; BC017658; AAH17658.1; -.
 DR Genew; HGNC:10730; SEMA4B.
 DR InterPro; IPR003659; Plexin-like.
 DR InterPro; IPR002165; Plexin_repeat.
 DR InterPro; IPR001627; Sema.
 DR Pfam; PF01437; PSI; 1.
 DR Pfam; PF01403; Sema; 1.
 DR SMART; SM00423; PSI; 1.
 DR SMART; SM00630; Sema; 1.
 DR Transmembrane; Immunoglobulin domain; Multigene family; Neurogenesis;
 KM Developmental protein; Glycoprotein; Signal; Polymorphism.
 FT SIGNAL 1 38
 FT CHAIN 39 832
 FT DOMAIN 39 712
 FT TRANSMEM 713 733
 FT DOMAIN 734 832
 FT DOMAIN 65 502
 FT DOMAIN 520 574
 FT DOMAIN 599 658
 FT DOMAIN 753 776
 FT DISULFID 606 651
 FT CARBOHYD 64 64
 FT CARBOHYD 91 91
 FT CARBOHYD 160 160
 FT CARBOHYD 405 405
 FT CARBOHYD 520 520
 FT CARBOHYD 625 625
 FT VARIANT 792 792
 FT CONFLICT 160 160
 FT CONFLICT 565 565
 SQ SEQUENCE 832 AA; 92192 MW; 29D58C1DD5E1C6B CRC64;
 Query Match 50.0%; Score 46; DB 1; Length 832;
 Best Local Similarity 61.5%; Pred. No. 9;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 VYOHOKAMPWIO 13
 Db 550 LHPQULATRPWIO 562
 RESULT 8
 PCTL_MOUSE
 ID PCTL_MOUSE STANDARD; PRT; 291 AA.
 AC Q9JMD3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE PCTP-like protein (PCTP-L) (Star-related lipid transfer protein 10)
 DE (STARD10) (START domain-containing protein 10) (Serologically defined
 DE colon cancer antigen 28 homolog)
 GN STARD10 OR SDCAG28 OR SDCAG28 OR PCTPL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=20297013; PubMed=10819773;
 RA Yamataka M., Koga M., Tanaka H., Nakamura Y., Ohta H., Yomogida K.,
 RA Tsuchida J., Iguchi N., Nojima H., Nozaki M., Matsunaga K.,
 RA Ouyama A., Tashimori K., Nishimune Y.;
 RT "Molecular cloning and characterization of phosphatidylcholine
 RT transfer protein-like protein gene expressed in murine haploid germ
 RT cells."
 RL Biol. Reprod. 62:1694-1701(2000).
 CC -1- FUNCTION: May play specific roles in sperm maturation or
 CC fertilization.
 CC -1- SUBCELLULAR LOCATION: In testis was predominantly detected at the
 CC flagella of elongated spermatozoa, with a strong signal also found

CC at the tail of epididymal sperm.
 CC -1- TISSUE SPECIFICITY: Testis, kidney, liver, and intestine with the
 CC highest level in the testis.
 CC -1- DEVELOPMENTAL STAGE: During male germ cell development, it was
 CC detected first in the 23-day-old mouse testis, and the signal
 CC increased with age.
 CC -1- SIMILARITY: Contains 1 START domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB031550; BAA92233.1; -.
 DR MGD; MG1:1660093; Stard10.
 DR InterPro; IPR002913; START.
 DR Pfam; PF01852; START; 1.
 DR SMART; SM00234; START; 1.
 DR PROSITE; PS50848; START; 1.
 FT DOMAIN 14 224
 FT SIGNAL 14 224
 SQ SEQUENCE 291 AA; 32951 MW; 598467C430D97080 CRC64;
 Query Match 47.8%; Score 44; DB 1; Length 291;
 Best Local Similarity 46.2%; Pred. No. 6;
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 3 OHOKAMPWIOPK 15
 Db 228 KHQPHKFWLHPE 240
 RESULT 9
 CHS1_LYCSES
 ID CHS1_LYCSES STANDARD; PRT; 389 AA.
 AC P23418;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1).
 DE CHS1.
 GN Lycopersicon esculentum (Tomato).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Solanum.
 OC NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cotyledon, Hypocotyl, and Leaf;
 RX MEDLINE=91117196; PubMed=1960524;
 RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
 RT "Molecular genetic analysis of chalcone synthase in Lycopersicon
 RT esculentum and an anocyanin-deficient mutant."
 RL Mol. Genet. 224:279-288(1990).
 CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: X55194; CA38980.1;
CC DR InterPro: IPR001099; N-C_synthase.
CC DR Pfam: PF00195; Chal_stll_synth_1.
CC DR Pfam: PF02797; Chal_stll_synth_1.
CC DR ProDom: PD000453; N-C_synthase; 1.
CC DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
CC KM Flavonoid biosynthesis; Transferase; Acyltransferase;
CC Multigene family.
CC FT ACT SITE 164 BY SIMILARITY.
CC SQ SEQUENCE 389 AA; 42852 MW; 553DC695EA96A8B CRC64;

QY Query Match 47.8%; Score 44; DB 1; Length 389;
QY Best Local Similarity 66.7%; Pred. No. 8.8;
QY Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 5 OKAMKPMIOPT 16
111 OKAIKMGQPKS 122

RESULT 10
CHS2_LYCES STANDARD; PRT; 389 AA.
ID CHS2_LYCES
AC P23419;
AT 01-NOV-1991 (Rel. 20, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
GN CHS2.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OC NCBI_TaxID=4081;
OX [1]
RN RP SEQUENCE FROM N.A.
RC TISSUE=Cotyledon, Hypocotyl, and Leaf;
RX MEDLINE=91117196; PubMed=1980524;
RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
RT "Molecular genetic analysis of chalcone synthase in Lycopersicon
RT esculentum and an anthocyanin-deficient mutant."
RT Mol. Gen. Genet. 224:279-288(1996).
CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: X55195; CA38981.1;
CC DR InterPro: IPR001099; N-C_synthase.
CC DR Pfam: PF00195; Chal_stll_synth_1.
CC DR Pfam: PF02797; Chal_stll_synth_1.
CC DR ProDom: PD000453; N-C_synthase; 1.
CC DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
CC KM Flavonoid biosynthesis; Transferase; Acyltransferase;
CC Multigene family.
CC FT ACT SITE 164 BY SIMILARITY.
CC SQ SEQUENCE 389 AA; 42730 MW; P92EA6BB3B5FC32F CRC64;

```

```

QY Query Match 47.8%; Score 44; DB 1; Length 389;
QY Best Local Similarity 66.7%; Pred. No. 8.8;
QY Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 5 OKAMKPMIOPT 16
111 OKAIKMGQPKS 122

RESULT 11
CHS2_SOLTU STANDARD; PRT; 389 AA.
ID CHS2_SOLTU
AC Q43188;
AT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
GN CHS2.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OC NCBI_TaxID=4113;
OX [1]
RN RP SEQUENCE FROM N.A.
RC STRAIN=cv. Red Pontiac;
RX MEDLINE=97141614; PubMed=8987872;
RA Jeon J.-H., Kim H.S., Choi K.H., Young Y.H., Young H., Byun S.-M.;
RT "Cloning and characterization of one member of the chalcone synthase
RT gene family from Solanum tuberosum L."
RT Biosci. Biotechnol. Biochem. 60:1907-1910(1996).
CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: U47738; AAB05239.1;
CC DR PIR: JCS136; JCS136.
CC DR InterPro: IPR001099; N-C_synthase.
CC DR Pfam: PF00195; Chal_stll_synth_1.
CC DR Pfam: PF02797; Chal_stll_synth_1.
CC DR ProDom: PD000453; N-C_synthase; 1.
CC DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
CC KM Flavonoid biosynthesis; Transferase; Acyltransferase;
CC Multigene family.
CC FT ACT SITE 164 BY SIMILARITY.
CC SQ SEQUENCE 389 AA; 42476 MW; 41618F9944958603 CRC64;

QY Query Match 47.8%; Score 44; DB 1; Length 389;
QY Best Local Similarity 66.7%; Pred. No. 8.8;
QY Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 5 OKAMKPMIOPT 16
111 OKAIKMGQPKS 122

RESULT 12

```

CHSA_PETHY STANDARD; PRT; 389 AA.

ID CHSA_PETHY
AC P08894;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase A (EC 2.3.1.74) (Naringenin-chalcone synthase A).
GN CHSA.

OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Petunia.
OC NCBI_TaxID=4102;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Violet 30;
RX MEDLINE=86286540; PubMed=3016642;
RA Koes R.E., Spelt C.E., Reif H.J., van den Elzen P.J.M., Veltkamp E.,
RA Mol J.N.M.;
RT "Floral tissue of Petunia hybrida (V30) expresses only one member of
RT the chalcone synthase multigene family.";
RL Nucleic Acids Res. 14:5229-5239(1986).
RN [2]

RP SEQUENCE FROM N.A.
RX STRAIN=cv. Violet 30;
RA MEDLINE=90034197; PubMed=2806915;
RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;
RT "Cloning and molecular characterization of the chalcone synthase
RT multigene family of Petunia hybrida.";
RL Gene 81:245-257(1989).
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- TISSUE SPECIFICITY: MAJOR EXPRESSED MEMBER OF THE GENE FAMILY IN
CC VARIOUS FLORAL TISSUES AND IN SEEDLINGS TREATED WITH UV LIGHT. IT
CC IS RELATIVELY LOW EXPRESSED IN TISSUE CULTURE MATERIAL.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04080; CA227718.1; -;
DR EMBL; X14591; CA22731.1; -;
DR PIR; A23643; SYPJCA.
DR PIR; J50308; SYPJCA.
DR InterPro; IPR001099; N-C synthase.
DR Pfam; PF00195; Chal_stc1_syn; 1.
DR Pfam; PF02797; Chal_stc1_syn; 1.
DR ProDom; PD000453; N-C synthase; 1.
DR PROSITE; PS00441; CHALCONE SYNTH; 1.
KM Flavonoid biosynthesis; Transferase; Acyltransferase;
KM Multigene family.
FT ACT SITE 164
FT ACT SITE 128
FT ACT SITE 128
SQ SEQUENCE 389 AA; 42525 MW; 0221B8DA1B96EAP6 CRC64;

Query Match 47.8%; Score 44; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 8.8;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 5 QKAKKWPQPT 16
|||||

Db 111 QKAKKWPQPKS 122

RESULT 13

CHSA_SOLTU STANDARD; PRT; 389 AA.

ID CHSA_SOLTU
AC Q41436;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase 1A (EC 2.3.1.74) (Naringenin-chalcone synthase 1A).
GN CHS1A.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.
OC NCBI_TaxID=4113;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Red Pontiac;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Young Y.-H., Young H., Byn S.-M.;
RT "Characterization of two members of the chalcone synthase gene family
RT from Solanum tuberosum L.";
RL (in) Plant Gene Register PGR96-027.
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U47739; AAB67734.1; -;
DR InterPro; IPR001099; N-C synthase.
DR Pfam; PF00195; Chal_stc1_syn; 1.
DR Pfam; PF02797; Chal_stc1_syn; 1.
DR ProDom; PD000453; N-C synthase; 1.
DR PROSITE; PS00441; CHALCONE SYNTH; 1.
KM Flavonoid biosynthesis; Transferase; Acyltransferase;
KM Multigene family.
FT ACT SITE 164
FT ACT SITE 164
SQ SEQUENCE 389 AA; 42562 MW; B181D9C6B9170A34 CRC64;

Query Match 47.8%; Score 44; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 8.8;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 5 QKAKKWPQPT 16
|||||

Db 111 QKAKKWPQPKS 122

RESULT 14

CHSB_SOLTU STANDARD; PRT; 389 AA.

ID CHSB_SOLTU
AC Q43163;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase 1B (EC 2.3.1.74) (Naringenin-chalcone synthase 1B).
GN CHS1B.
OS Solanum tuberosum (Potato).

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv, Red Pontiac;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Young Y.-H., Young H., Byun S.-M.;
RT "Characterization of two members of the chalcone synthase gene family
RT from Solanum tuberosum L.";
RL (in) Plant Gene Register PCR96-027.
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U47740; AAB67735.1; -.
DR InterPro; IPR001099; N-C_synthase.
DR Pfam; PF00195; Chal_still_synth; 1.
DR Pfam; PF02797; Chal_still_synth; 1.
DR ProDom; PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KM Multigene family.
FT ACT SITE 164
SQ SEQUENCE 389 AA; 42548 MW; E7334A1A34D8D1CC CRC64;

Query Match 47.8%; Score 44; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 8.8;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKT 16
DB 111 OKAIKMGQPKS 122

RESULT 15
CHSE_IPOPU
ID CHSE_IPOPU STANDARD; PRT; 389 AA.
AC 022047;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase E (EC 2.3.1.74) (Naringenin-chalcone synthase E)
DE (CHS-E).
GN CHSE.
OS Ipomoea purpurea (Common morning glory).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=4121;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv, FR-35; TISSUE=flower buds;
RX MEDLINE=97393496; PubMed=924990;
RA Fukada-Tanaka S., Hoshino A., Hiatomi Y., Habu Y., Hasebe M.,
RA Iida S.;
RT "Identification of new chalcone synthase genes for flower pigmentation
RT in the Japanese and common morning glories.";

```

```

RL Plant Cell Physiol. 38:754-758(1997).
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB001827; BAA21789.1; -.
DR PIR; T07799; T07799.
DR InterPro; IPR001099; N-C_synthase.
DR Pfam; PF00195; Chal_still_synth; 1.
DR Pfam; PF02797; Chal_still_synth; 1.
DR ProDom; PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KM Multigene family.
FT ACT SITE 164
SQ SEQUENCE 389 AA; 42677 MW; 0CC81302CC8A56A CRC64;

Query Match 47.8%; Score 44; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 8.8;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKT 16
DB 111 OKAIKMGQPKS 122

Search completed: July 30, 2003, 16:24:28
Job time : 6.2381 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 10.1905 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-1

Perfect score: 92
Sequence: 1 VYQHOKMKRWIQPKT 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_76:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	232	1 KABS02	alpha-s2-casein pr
2	77	83.7	223	1 KASHS2	alpha-s2-casein pr
3	77	83.7	223	2 JN0547	alpha-s2-casein pr
4	77	83.7	223	2 I46995	alpha-s2-casein C
5	56	60.9	235	2 A48383	alpha-s2-casein -
6	48	52.2	418	2 F72370	phoH-related prote
7	46.5	50.5	363	2 T29156	hypothetical prote
8	46.5	50.5	364	2 T29153	hypothetical prote
9	46	50.0	180	2 S39775	alpha-s2-casein fo
10	46	50.0	265	2 B58930	succinate dehydrog
11	45	48.9	311	2 T15997	hypothetical prote
12	44	47.8	319	2 S18136	naringenin-chalcon
13	44	47.8	328	2 H75008	naringenin-chalcon
14	44	47.8	340	2 T10957	hypothetical prote
15	44	47.8	389	1 SYPUJ	naringenin-chalcon
16	44	47.8	389	1 SYPUJ	naringenin-chalcon
17	44	47.8	389	1 SYPUJ	naringenin-chalcon
18	44	47.8	389	1 JCS136	naringenin-chalcon
19	44	47.8	389	2 T07799	naringenin-chalcon
20	44	47.8	389	2 T07799	naringenin-chalcon
21	44	47.8	389	2 T07799	naringenin-chalcon
22	44	47.8	389	2 T07799	naringenin-chalcon
23	44	47.8	389	2 T07799	naringenin-chalcon
24	44	47.8	389	2 T07799	naringenin-chalcon
25	44	47.8	389	2 T07799	naringenin-chalcon
26	44	47.8	389	2 T07799	naringenin-chalcon
27	44	47.8	389	2 T07799	naringenin-chalcon
28	44	47.8	389	2 T07799	naringenin-chalcon
29	44	47.8	389	2 T07799	naringenin-chalcon

30	42.5	46.2	690	1 OYRTA1	guanylate cyclase
31	42	45.7	143	1 KEMS	epsilon-casein pre
32	42	45.7	186	2 S03626	neurogenic gene co
33	42	45.7	217	2 T04292	hypothetical prote
34	42	45.7	224	2 D46177	enhancer of split
35	42	45.7	248	2 T01880	hypothetical prote
36	42	45.7	327	2 C71146	hypothetical prote
37	42	45.7	389	1 S36653	kpsd protein - Esc
38	42	45.7	404	1 S62440	mevalonate kinase
39	42	45.7	589	2 I37187	arylsulfatase B (B
40	42	45.7	1091	2 T34107	hypothetical prote
41	41	44.6	232	2 B97147	enzyme of sideroph
42	41	44.6	324	2 S58061	amelanin 2 - rat
43	41	44.6	327	1 S75716	molybdenum cofacto
44	41	44.6	367	2 T52363	hypothetical prote
45	41	44.6	398	2 S16275	naringenin-chalcon

ALIGNMENTS

RESULT 1

KABS02
alpha-s2-casein precursor - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 01-Sep-1981 #sequence revision 03-Feb-1994 #ext_change 22-Jun-1999
C/Accession: J02008; A29087; A91438; S66626; A03107
R/Groenen, M.A.M.; Dijkhof, R.J.M.; Verstege, A.J.M.; van der Poel, J.J.
Gene 123, 187-193, 1993
A/Title: The complete sequence of the gene encoding bovine alpha-s2-casein.
A/Reference number: J02008; MUID:93154583; PMID:8428658
A/Accession: J02008
A/Status: translation not shown
A/Molecule type: DNA
A/Residues: 1-222 <GRO>
A/Cross-references: GB:M94327
R/Stewart, A.F.; Bomsing, J.; Beattie, C.W.; Shah, F.; Willis, I.M.; Mackinlay, A.G.
Mol. Biol. Evol. 4, 231-241, 1987
A/Title: Complete nucleotide sequences of bovine alpha-s2- and beta-casein cDNAs: compar
A/Reference number: A93062; MUID:18818989; PMID:2833669
A/Accession: A29087
A/Status: translation not shown
A/Molecule type: mRNA
A/Residues: 1-222 <STB>
R/Brignon, G.; Ribadeau Dumas, B.; Mercier, J.C.; Pelissier, J.P.; Das, B.C.
FEBS Lett. 76, 274-279, 1977
A/Title: Complete amino acid sequence of bovine alpha-S2-casein.
A/Reference number: A91438; MUID:77185633; PMID:862906
A/Accession: A91438
A/Molecule type: protein
A/Residues: 16-101, 104-222 <BRI>
A/Note: Four fractions, previously designated s2, s3, s4, and s6, appear to have the sa
these
R/Groenclau, F.; Joudrier, P.; Mahe, M.F.
J. Dairy Res. 46, 211-213, 1979
A/Title: A genetic and biochemical analysis of a polymorphism of bovine alpha-S2-casein
A/Reference number: A92771; MUID:79239837; PMID:469044
A/Content: annotation; D allele
A/Note: the sequence of the D allele has a deletion of nine residues, which may be 49-58
R/Zucht, H.D.; Raide, M.; Ademann, K.; Maegert, H.U.; Forsemann, W.G.
FEBS Lett. 372, 185-188, 1995
A/Title: Casocidin-I: a casein-alpha(s2) derived peptide exhibits antibacterial activity
A/Reference number: S66626; MUID:9600204; PMID:755666
A/Accession: S66626
A/Molecule type: protein
A/Residues: 165-203 <ZUC>
C/Comment: The sequence of the A allele is shown.
C/genetic:
A/gene: alphaS2ca
A/map position: 6
A/introns: 17/3; 26/3; 33/3; 47/3; 56/3; 65/3; 74/3; 82/3; 97/3; 138/3; 147/3; 156/3; 161

C:Superfamily: alpha-s2-casein
 C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-223/Product: alpha-s2-casein #status experimental <MAT>
 F:23,24,25,31,71,72,73,76,144,146,158/Binding site: phosphate (Ser) (covalent) #status F

Query Match 100.0%; Score 92; DB 1; Length 223;
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKT 16
 Db 199 VQHOKAMKPMIOPKT 213

RESULT 2

KASHS2
 alpha-s2-casein precursor - sheep
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C>Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 22-Jun-1999
 C:Accession: A25070; S17856
 R:Boismard, M.; Petrisant, G.
 Biochimie 67, 1043-1051, 1985
 A>Title: Complete sequence of ovine alpha-s2-casein messenger RNA.
 A:Reference number: A25070; MUID:86104467; PMID:3002499
 A:Accession: A25070
 A:Molecule type: mRNA
 A:Residues: 1-223 <BO1>

A:Cross-references: GB:X03238; NID:g1238; PIDN:CAA6993.1; PID:g732894
 A>Note: 64 Asn was also found
 R:Boismard, M.; Hue, D.; Bouniol, C.; Mercier, J.C.; Gaye, P.
 Eur. J. Biochem. 201, 633-641, 1991
 A>Title: Multiple mRNA species code for two non-allelic forms of ovine alpha-s2-casein.
 A:Reference number: S17856; MUID:92037619; PMID:1935959
 A:Accession: S17856

A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-93, 'T', 95-223 <BO2>
 C:Superfamily: alpha-s2-casein
 C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-223/Product: alpha-s2-casein #status predicted <KA2>
 F:23,24,25,32,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status F:53,88,146,154,170,198/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 83.7%; Score 77; DB 1; Length 223;
 Best Local Similarity 87.5%; Pred. No. 4.6e-05;
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKT 16
 Db 199 VQHOKAMKPMIOPKT 214

RESULT 3

JN0547
 alpha-s2-casein precursor - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C>Date: 31-Dec-1993 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: S33861; S33860; JN0547; S20620
 R:Bouniol, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Protein Seq. Data Anal. 5, 213-218, 1993
 A>Title: Characterization of goat allelic alpha-s2-caseins A and B: further evidence of

A:Reference number: S33860
 A:Accession: S33861
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-223 <BOU>
 A>Note: allele B
 A:Accession: S33860
 A:Molecule type: protein
 A:Residues: 16-78, 'E', 80-223 <BO2>
 A>Note: sequence deduced from compositional analysis of peptides

R:Bouniol, C.
 Gene 125, 235-236, 1993
 A>Title: Sequence of the goat alpha-s2-casein-encoding cDNA.
 A:Reference number: JN0547; MUID:93216130; PMID:8462880
 A:Accession: JN0547

A:Molecule type: mRNA
 A:Residues: 1-78, 'E', 80-223 <BO3>
 A:Cross-references: EMBL:X65160; NID:g9955; PIDN:CAA6278.1; PID:g9956
 A>Note: allele A
 C:Superfamily: alpha-s2-casein

C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-223/Product: alpha-s2-casein #status experimental <MAT>
 F:23,24,25,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status F:53,146/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 83.7%; Score 77; DB 2; Length 223;
 Best Local Similarity 87.5%; Pred. No. 4.6e-05;
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKT 16
 Db 199 VQHOKAMKPMIOPKT 214

RESULT 4

I46995
 alpha-s2-casein C - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C>Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 13-Aug-1999
 C:Accession: I46995
 R:Bouniol, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Anim. Genet. 25, 173-177, 1994
 A>Title: Biochemical and genetic analysis of variant C of caprine alpha-s2-casein (Capra
 A:Reference number: I46995; MUID:95030556; PMID:7943951
 A:Accession: I46995
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-223 <BOU>
 A:Cross-references: GB:S74171; NID:g707033; PIDN:AA83216.1; PID:g707034
 C:Superfamily: alpha-s2-casein

Query Match 83.7%; Score 77; DB 2; Length 223;
 Best Local Similarity 87.5%; Pred. No. 4.6e-05;
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKT 16
 Db 199 VQHOKAMKPMIOPKT 214

RESULT 5

A48383
 alpha-s2-casein - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
 C:Accession: A48383
 R:Alexander, L.J.; Das Gupta, N.A.; Beattie, C.W.
 Anim. Genet. 23, 365-367, 1992
 A>Title: The sequence of porcine alpha-s2-casein cDNA.
 A:Reference number: A48383; MUID:92367960; PMID:1503276
 A:Accession: A48383
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-235 <ALB>
 A:Experimental source: mammary gland
 A>Note: sequence inconsistent with the nucleotide translation
 A>Note: sequence extracted from NCBI backbone (NCBIN:110884; NCBIP:110885)
 C:Superfamily: alpha-s2-casein

Query Match 60.9%; Score 56; DB 2; Length 235;
 Best Local Similarity 68.8%; Pred. No. 0.12;
 Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHOKAMKRWIOPKT 16
|:|||||:
Db 208 VHOKOKAMKRWNIKT 223

RESULT 6
F72370
pHcH-related protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: F72370
R:Jelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: F72370
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-418 <ARN>
A:Cross-references: GB:AE001726; GB:AE000512; NID:g4980992; PIDN:AND35580.1; PID:g498100
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0495

Query Match 52.2%; Score 48; DB 2; Length 418;
Best Local Similarity 70.0%; Pred. No. 4.4;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 OKAMKRWIO 14
|:|||||:
Db 285 EKKMKRWIO 294

RESULT 7
T29156
hypothetical protein F47B3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29156
R:Du, Z.; Le, T.T.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid F47B3.
A:Reference number: Z20579
A:Accession: T29156
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-363 <DUZ>
A:Cross-references: EMBL:U97017; PIDN:AAB52356.1; GSPDB:GN00019; CESP:F47B3.1
C:Genetics:
A:Gene: CESP:F47B3.1
A:Map position: 1
A:Introns: 6/2; 54/3; 98/2; 251/3; 291/3; 316/3

Query Match 50.5%; Score 46.5; DB 2; Length 363;
Best Local Similarity 69.2%; Pred. No. 6.7;
Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

QY 4 HOKAM--KRWIO 13
|:|||||:
Db 48 HOKAMIVKRWIO 60

RESULT 8
T29153
hypothetical protein F47B3.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29153
R:Du, Z.; Le, T.T.

submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid F47B3.
A:Reference number: Z20579
A:Accession: T29153
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-364 <DUZ>
A:Cross-references: EMBL:U97017; PIDN:AAB52361.1; GSPDB:GN00019; CESP:F47B3.6
A:Experimental source: strain Bristol N2; clone F47B3
C:Genetics:
A:Gene: CESP:F47B3.6
A:Map position: 1
A:Introns: 7/2; 55/3; 99/2; 252/3; 292/3; 317/3

Query Match 50.5%; Score 46.5; DB 2; Length 364;
Best Local Similarity 69.2%; Pred. No. 6.7;
Matches 9; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

QY 4 HOKAM--KRWIO 13
|:|||||:
Db 49 HOKAMIVKRWIO 61

RESULT 9
S39775
alpha-s2-casein form a precursor - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 13-Aug-1999
C:Accession: S39775
R:Dawson, S.P.; Wilde, C.J.; Tighe, P.J.; Mayer, R.U.
Biochem. J. 296, 777-784, 1993
A:Title: Characterization of two novel casein transcripts in rabbit mammary gland.
A:Reference number: S39775; MUID:94107245; PMID:8280077
A:Accession: S39775
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-180 <DAW>
A:Cross-references: EMBL:X76907; NID:g439525; PIDN:CAA54228.1; PID:g439526
C:Superfamily: gamma-casein

Query Match 50.0%; Score 46; DB 2; Length 180;
Best Local Similarity 56.2%; Pred. No. 3.9;
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 VYOHOKAMKRWIOPKT 16
|:|||||:
Db 99 VTQOQIVMMPKRWIKT 114

RESULT 10
B58930
succinate dehydrogenase (ubiquinone) (BC 1.3.5.1) iron-sulfur protein - Cyanidioschyzon
C:Species: Cyanidioschyzon merolae
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 20-Jun-2000
C:Accession: B58930
R:Ohira, N.; Sato, N.; Kuroiwa, T.
Nucleic Acids Res. 26, 5190-5198, 1998
A:Title: Structure and organization of the mitochondrial genome of the unicellular red
A:Reference number: A58930; MUID:99030526; PMID:9801318
A:Accession: B58930
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-265 <ARN>
A:Cross-references: GB:D89861; NID:g4115781; PIDN:BAA34653.1; PID:g3927865
C:Genetics:
A:Gene: sdhB
A:Genome: mitochondrion
C:Superfamily: fumarate reductase iron-sulfur protein; ferredoxin 2[4Fe-4S] homology; fe
C:Keywords: 2Fe-2S; iron-sulfur protein; metalloprotein; mitochondrion; oxidoreductase
F;60-102/Domain: ferredoxin 2[4Fe-4S] homology <FER1>
F;165-247/Domain: ferredoxin 2[4Fe-4S] homology <FER2>
F;81,86,89,101/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

A;Map position: V
A;Note: chsJ is expressed in various floral tissues and UV illuminated seedlings
C;Superfamily: chalcone synthase
C;Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 47.8%; Score 44; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAKPPIOPKT 16
Db 111 OKAKKWKGPXS 122

Search completed: July 30, 2003, 16:31:14
Job time : 11.1905 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 31.5238 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-1
Perfect score: 92
Sequence: 1 VYQHOKAMKRWIQPKT 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq.19Jun03.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	16	21	AA198337
2	92	100.0	21	21	AA198338
3	92	100.0	24	15	AA1980481
4	92	100.0	24	21	AA198339
5	92	100.0	25	13	AA1980055
6	92	100.0	25	21	AA198340
7	92	100.0	31	18	AA1982219
8	92	100.0	31	23	AA19817466
9	92	100.0	222	18	AA19832220

10	92	100.0	222	23	AA19817468	Bovine alpha-S2 ca
11	81.5	88.6	26	20	AA1983885	Bitidobacterium bi
12	77	83.7	223	23	AA19817469	Alpha-S2 casein pr
13	77	83.7	223	23	AA19817470	Goat alpha-S2 case
14	77	83.7	223	23	AA19817471	Capra hircus alpha
15	77	83.7	223	23	AA19817473	Sheep alpha-S2 cas
16	77	83.7	223	23	AA19817475	Alpha-S2 casein pr
17	58	63.0	19	18	AA1982218	Alpha-S2 casein pr
18	58	63.0	19	23	AA19817465	Bovine alpha-S2 ca
19	56	60.9	234	23	AA19817474	Pig alpha-S2 casei
20	53	57.6	9	17	AA19804282	Milk derived anti-
21	49	53.3	8	17	AA19804278	Milk derived anti-
22	48	52.2	103	22	AA19804204	Human immune/haema
23	48	52.2	191	21	AA19834754	Human secreted pro
24	48	52.2	328	22	AA1982405	S. epidermidis ope
25	48	52.2	328	23	AA1983854	Staphylococcus epi
26	48	52.2	575	22	AA19827239	Human EXMAD-17 SEQ
27	48	52.2	704	22	AA1985234	Human protein sequ
28	48	52.2	704	23	AA19869565	Human polypeptide
29	46	50.0	315	22	AA19807420	Novel human extrac
30	46	50.0	372	23	AA19806109	Human NS protein s
31	46	50.0	675	23	AA19864686	Human protein SEQ
32	46	50.0	681	21	AA19801396	Neuron-associated
33	46	50.0	791	23	AA19807413	Human NOV7 protein
34	46	50.0	832	22	AA19803818	Human gene 1 encod
35	46	50.0	832	23	AA19864522	Human albumin fusi
36	46	50.0	837	21	AA19894410	Human PRO1480 (UNO
37	46	50.0	837	22	AA19829250	Human PRO polypept
38	46	50.0	837	22	AA19866159	Protein of the inv
39	46	50.0	837	23	AA19865100	Hypoxia-induced pr
40	46	50.0	837	24	AA19848242	Human bladder canc
41	46	50.0	837	24	AA198071338	Human albumin prot
42	46	50.0	837	24	AA19865795	Human secreted/cra
43	46	50.0	837	24	AA19865128	Novel human secret
44	46	50.0	837	24	AA19806732	Human secreted/cra
45	46	50.0	837	24	AA19865490	Human PRO polypept

ALIGNMENTS

RESULT 1
ID AAY88337
AAY88337 strand; peptide; 16 AA.
XX AAY88337;
AC AAY88337;
XX 14-JUL-2000 (first entry)
DT 14-JUL-2000 (first entry)
XX 14-JUL-2000 (first entry)
DE Internal peptide fragment of bovine alpha-S2 casein #1.
XX Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
XX antibacterial peptide; lactoferrin; antiviral; antitumour activity.
XX Bos sp.
XX WO200015655-A1.
PN WO200015655-A1.
XX 23-MAR-2000.
PD 23-MAR-2000.
XX 15-SEP-1999; 99WO-EP07002.
PF 15-SEP-1999; 99WO-EP07002.
XX 15-SEP-1998; 98EP-020107.
PR 08-JUN-1999; 98EP-0201815.
XX (NIZO-) NIZO FOOD RES.
PA (NIZO-) NIZO FOOD RES.
XX Vissers S, Recio I;
PI Vissers S, Recio I;
XX WPI; 2000-271377/23.
DR WPI; 2000-271377/23.
XX Novel process for producing peptides with e.g. antimicrobial activity
XX from biological fluids such as milk, whey or blood comprises contacting

PT fluid with chromatographic medium to adsorb peptide domain of interest
 PS Claim 14; Page 22; 41pp; English.

XX This sequence represents an internal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

XX Sequence 16 AA;

Query Match 100.0%; Score 92; DB 21; Length 16;
 Best Local Similarity 100.0%; Pred. No. 1.2e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VYOHOKAMKPMIOPKT 16
 |||||
 Db 1 VYOHOKAMKPMIOPKT 16

RESULT 2

AAV86338
 ID AAV86338 standard; peptide; 21 AA.

XX AAV86338;

DT 14-JUN-2000 (first entry)

DE Internal peptide fragment of bovine alpha-S2 casein #2.

XX Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
 KM antibacterial peptide; lactoferrin; antiviral; antitumour activity.

XX Bos sp.

XX WO200015655-A1.

XX 23-MAR-2000.

PF 15-SEP-1999; 99WO-EP07002.

XX 15-SEP-1998; 98EP-0203107.

PR 08-JUN-1999; 99EP-0201815.

XX (NIZO-) NIZO FOOD RES.

PA Viasser S, Recio I;

XX WPI; 2000-271377/23.

XX Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest

XX Claim 14; Page 22; 41pp; English.

XX This sequence represents an internal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The

CC Invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

XX Sequence 21 AA;

Query Match 100.0%; Score 92; DB 21; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.6e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VYOHOKAMKPMIOPKT 16
 |||||
 Db 1 VYOHOKAMKPMIOPKT 16

RESULT 3

AA60481
 ID AA60481 standard; peptide; 24 AA.

XX AA60481;

DT 19-APR-1995 (first entry)

DE Peptide used in tranquilliser.

XX Peptide; tranquilliser; food; foodstuff; mental stabilisation.

XX Synthetic.

XX JP06211689-A.

XX 02-AUG-1994.

PF 19-JAN-1993; 93JP-0024811.

PR 19-JAN-1993; 93JP-0024811.

XX (KANE) KANEBO LTD.

XX WPI; 1994-283276/35.

XX Synthetic peptide(s) used in tranquilliser - also used in the
 PT production of a food for mental stabilisation.

XX Claim 3; Page 1; 5pp; Japanese.

XX The peptide may be used as a component of a tranquilliser which may
 CC be administered orally and is low in toxicity. The peptide may also
 CC be used as a component of a foodstuff which may be used for mental
 CC stabilisation. See also AA60480.

XX Sequence 24 AA;

Query Match 100.0%; Score 92; DB 15; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VYOHOKAMKPMIOPKT 16
 |||||
 Db 1 VYOHOKAMKPMIOPKT 16

RESULT 4

AAV88339
 ID AAV88339 standard; peptide; 24 AA.
 XX
 AC AAV88339;
 XX
 DT 14-JUL-2000 (first entry)
 XX
 DE Internal peptide fragment of bovine alpha-S2 caesin #3.
 XX
 KM Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;
 KM antibacterial peptide; lactoferrin; antiviral; antitumour activity.
 XX
 OS Bos sp.
 XX
 PN WO200015655-A1.
 XX
 PD 23-MAR-2000.
 XX
 PF 15-SEP-1999; 99WO-EP07002.
 XX
 PR 15-SEP-1998; 98EP-0203107.
 PR 08-JUN-1999; 99EP-0201815.
 XX
 PA (NIZO-) NIZO FOOD RES.
 XX
 PI Visser S, Recto I;
 XX
 DR WPI; 2000-271377/23.
 XX
 PT Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest
 PT
 PT
 PS Claim 14; Page 22; 41pp; English.
 XX
 XX
 CC This sequence represents an internal fragment of bovine alpha-S2 caesin
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.
 CC
 SQ Sequence 24 AA;
 XX
 XX
 Query Match 100.0%; Score 92; DB 21; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKFWIQPKT 16
 DB 1 VYQHOKAMKFWIQPKT 16
 XX
 XX
 RESULT 5
 ID AAR20055 standard; peptide; 25 AA.
 XX
 AC AAR20055;
 XX
 DT 25-MAR-2003 (updated)
 DT 26-MAR-1992 (first entry)
 PT

DE Casein peptide.
 XX
 KM Casein; platelet; aggregation; thrombosis; collagen.
 XX
 OS Synthetic.
 XX
 PN JP03255095-A.
 XX
 PD 13-NOV-1991.
 XX
 PF 02-MAR-1990; 90JP-0052554.
 XX
 PR 02-MAR-1990; 90JP-0052554.
 XX
 PA (KANE) KANEBO LTD.
 XX
 DR WPI; 1992-002669/01.
 XX
 PT Casein peptide(s) for treating thrombosis - as inhibitors of
 PT platelet aggregation caused by adenosine-5'-di-phosphate and
 PT collagen and as biochemical reagents
 XX
 PS Claim 1; Page 1; 8pp; Japanese.
 XX
 CC Leu25 may be omitted. The peptide and its salts have inhibiting
 CC activity against platelet aggregation caused by adenosine-5'-
 CC di-phosphate and by collagen. They are useful for prevention and
 CC treatment of thrombosis and are also useful as a biochemical reagent.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 CC
 SQ Sequence 25 AA;
 XX
 XX
 Query Match 100.0%; Score 92; DB 13; Length 25;
 Best Local Similarity 100.0%; Pred. No. 2e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKFWIQPKT 16
 DB 1 VYQHOKAMKFWIQPKT 16
 XX
 XX
 RESULT 6
 ID AAY88340 standard; peptide; 25 AA.
 XX
 AC AAY88340;
 XX
 DT 14-JUL-2000 (first entry)
 XX
 DE C-terminal peptide fragment of bovine alpha-S2 caesin.
 XX
 KM Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;
 KM antibacterial peptide; lactoferrin; antiviral; antitumour activity.
 XX
 OS Bos sp.
 XX
 PN WO200015655-A1.
 XX
 PD 23-MAR-2000.
 XX
 PF 15-SEP-1999; 99WO-EP07002.
 XX
 PR 15-SEP-1998; 98EP-0203107.
 PR 08-JUN-1999; 99EP-0201815.
 XX
 PA (NIZO-) NIZO FOOD RES.
 XX
 PI Visser S, Recto I;
 XX
 DR WPI; 2000-271377/23.
 XX
 PT Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting

PT fluid with chromatographic medium to adsorb peptide domain of interest

PT -

XX Claim 14, Page 22, 41pp; English.

PS

XX This sequence represents a C-terminal fragment of bovine alpha-S2 casein

CC protein. The peptide is an example of a peptide with antibacterial

CC activity that can be produced by the process of the invention. The

CC invention relates to a process for producing peptides from biological

CC fluids. The process comprises chromatography of the biological fluid, in

CC situ hydrolysis of selectively bound peptides, washing to remove unbound

CC peptide, and elution of the peptides of interest. The process is used for

CC producing peptides from biological fluids, such as milk, whey or blood.

CC For example, the process can be used to produce antibacterial peptides

CC derived from lactoferrin, using cheese whey as a starting material. The

CC peptides obtained have preferably antimicrobial and/or antiviral and/or

CC antitumour activity. The process of the invention is relatively simple

CC and generally economically and technically more attractive than those

CC methods previously used. The method provides high yield peptides with a

CC selected activity of interest without the need for intermediate

CC purification of the precursor protein.

XX

SQ Sequence 25 AA;

Query March 100.0%; Score 92; DB 21; Length 25;

Best Local Similarity 100.0%; Pred. No. 2e-07; Mismatches 0; Gaps 0;

Matches 16; Conservative 0; Indels 0;

QY 1 VYQHOKAMKPMIQTPT 16

Db 1 VYQHOKAMKPMIQTPT 16

RESULT 7

AAW32219

ID AAW32219 standard; peptide; 31 AA.

XX AAW32219;

AC

XX 03-FEB-1998 (first entry)

DT

XX

DE Alpha-S2 casein precursor C-terminal peptide fragment #5.

KW Alpha-S2 casein precursor; growth promoting; mitogenic assay;

KM platelet-derived growth factor; insulin-like derived growth factor;

XX Synthetic.

OS Bos taurus.

XX

PN MO9716460-A1.

PD

XX 09-MAY-1997.

PE 31-OCT-1996; 96MO-GB02658.

PR 31-OCT-1995; 95GB-0022302.

XX (UYLI-) UNIV LIVERPOOL.

PA

XX

PI Liu Q, Smith JA, Wilkinson MC;

DR WPI, 1997-272048/24.

XX

PT Manufacture of medicament or foodstuff for promoting growth - using

PT peptide(s) with a sequence identical to the C-terminal end of an

PT alpha-S2 casein precursor

PS Claim 12; Page 21; 33pp; English.

XX

CC The present sequence, which is substantially identical to the C-terminal

CC end of an alpha-S2 casein precursor, was found after storage in PBS to

CC exhibit growth promoting activity for rat mammary fibroblast cell line

CC (Rama 27), which is not significantly stimulated by platelet-derived

[illegible]

Db 7 VYOHOKAMKFWIQPKT 22

RESULT 9

AAW32220 standard; protein; 222 AA.

AAW32220;

03-FEB-1998 (first entry)

Bovine alpha-S2 casein precursor.

Bovine alpha-S2 casein precursor; growth promoting; mitogenic assay;

platelet-derived growth factor; insulin-like derived growth factor;

Bos taurus.

WO9716460-A1.

09-MAY-1997.

31-OCT-1996; 96WO-GB02658.

31-OCT-1995; 95GB-0022302.

(UYLF-) UNIV LIVERPOOL.

Liu Q, Smith JA, Wilkinson MC;

WPI; 1997-272048/24.

Manufacture of medicament or foodstuff for promoting growth - using

peptide(s) with a sequence identical to the C-terminal end of an

alpha-S2 casein precursor

Disclosure; Page 3; 33pp; English.

The present sequence represents bovine alpha-S2 casein precursor.

Peptides having an amino acid sequence which is substantially identical

to the C-terminal end of an alpha-S2 casein precursor, are used for

manufacture of a medicament or foodstuff for promoting growth in humans

or animals.

Sequence 222 AA;

Query Match 100.0%; Score 92; DB 18; Length 222;

Best local Similarity 100.0%; Pred. No. 2.1e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VYOHOKAMKFWIQPKT 16

198 VYOHOKAMKFWIQPKT 213

AAE17468 standard; Protein; 222 AA.

AAE17468;

22-APR-2002 (first entry)

Bovine alpha-S2 casein precursor protein.

Bovine; alpha-S2 casein; dermatological; anti-inflammatory; fibroblast;

collagen; keratinocyte; skin regeneration; medication; toothpaste; aging;

chewing gum; cosmetic; wrinkling; periodontal disease.

Bos sp.

WO200202133-A2.

10-JAN-2002.

13-JUN-2001; 2001WO-GB02601.

30-JUN-2000; 2000GB-0016189.

(PEPS-) PEPSYN LTD.

Smith JA;

WPI; 2002-154690/20.

Use of peptide or its derivative containing an amino acid sequence in

alpha-S2 casein precursor in the manufacture of a medicament for

alleviating or preventing periodontal disease and an effect of aging in

skin -

Claim 8; Page 6; 27pp; English.

The invention relates to a composition comprising a peptide or its

derivative. The peptide contains an amino acid sequence from alpha-S2

casein precursor. The peptides stimulate the growth of fibroblasts, and

thus the synthesis and secretion of collagen. The peptides also stimulate

the growth of keratinocytes, which aid in the formation and regeneration

of skin surface. The peptide is useful in the manufacture of a medicament

in the form of a toothpaste or a chewing gum, for alleviating or

preventing periodontal disease and a medicament in the form of a cosmetic

composition for alleviating or preventing an effect of aging,

particularly wrinkling of the skin. The present sequence is bovine

alpha-S2 casein precursor protein.

Sequence 222 AA;

Query Match 100.0%; Score 92; DB 23; Length 222;

Best local Similarity 100.0%; Pred. No. 2.1e-06;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VYOHOKAMKFWIQPKT 16

198 VYOHOKAMKFWIQPKT 213

AAW93885 standard; peptide; 26 AA.

AAW93885;

25-JUN-1999 (first entry)

Bifidobacterium bifidus stimulating peptide 21.

Bifidogenic peptide; protease; treatment; microbe-related disease;

bacteria; fungi; yeast; protozoa; virus; mycoplasma; filaria; plasmodia;

infection; inflammation; microbial induced tumour; degenerative disorder;

diarrhoea; colic; oral microflora; intestinal microflora; caries;

vaginal microflora.

Bifidobacterium bifidus.

WO9914231-A2.

25-MAR-1999.

16-SEP-1998; 98WO-EP05899.

11-FEB-1998; 98DE-1005385.

16-SEP-1997; 97DB-1040604.

(FORS/) FORSSMANN W.

Forsmann W, Liepke C, Zucht H;

XX WPI; 1995-244022/20.
 DR
 XX
 PT Milk-derived peptides that stimulate Bifidobacterium bifidus
 XX
 PS Claim 2; Page 3; 25pp; German.
 CC This invention describes milk-derived bifidogenic peptides and their
 CC active derivatives or fragments, and combinations of them produced by
 CC chemical coupling. Such are produced from bovine or human milk by
 CC treatment for 2 hr with proteases, then centrifuging to remove fat and
 CC acidifying to pH 2 to precipitate proteins. The solution phase is then
 CC subjected to reverse-phase high-performance liquid chromatography (HPLC)
 CC and cation-exchange HPLC, the fractions adjusted to salt content below
 CC 25 mM (by dialysis or reverse-phase HPLC) and tested for activity by
 CC growing Bifidobacterium bifidus and Escherichia coli in presence of the
 CC fractions. Those fractions for which (Bw-B0)-(Bw-F0) is at least 0.15
 CC are selected where Bw = germ count after 16 hr culture of B. bifidus in
 CC 50% Elliker broth containing peptide at 0.2 mg/mL, B0 = germ count under
 CC similar conditions in a peptide-free control, Bw = germ count after 16
 CC hr culture of E. coli in 3 g/l tryptic broth containing peptide at
 CC 0.2 mg/mL, B0 = germ count under similar conditions in a peptide-free
 CC control. The peptides AAW3865-W3888 are used to treat microbe-related
 CC diseases caused by bacteria, fungi, yeast, protozoa, viruses,
 CC mycoplasma, filaria and plasmodia, e.g. infections, inflammation,
 CC microbially induced tumours or degenerative disorders, diarrhoea, colic,
 CC abnormalities in oral, intestinal or vaginal microflora, or caries.
 XX
 SQ Sequence 26 AA;
 Query Match 88.6%; Score 81.5; DB 20; Length 26;
 Best Local Similarity 94.1%; Pred. No. 8.8e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 VYOHQKAM-KPWIOPT 16
 DB 1 VYOHQKAMKPWIOPT 17
 RESULT 12
 AAE17469
 ID AAE17469 standard; Protein; 223 AA.
 AC AAE17469;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Alpha-S2 casein precursor (alpha-S2-CN) protein #1.
 XX
 KW Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
 KW fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
 KW toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Unidentified.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin

XX Disclosure; Page 8; 27pp; English.
 PS
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is alpha-S2
 CC casein precursor (alpha-S2-CN) protein.
 XX
 SQ Sequence 223 AA;
 Query Match 83.7%; Score 77; DB 23; Length 223;
 Best Local Similarity 87.5%; Pred. No. 0.00045;
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 VYOHQKAMKPWIOPT 16
 DB 199 VDHQKAMKPWIOPT 214
 RESULT 13
 AAE17470
 ID AAE17470 standard; Protein; 223 AA.
 AC AAE17470;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Goat alpha-S2 casein E precursor protein.
 XX
 KW Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KW chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin
 XX
 PS Claim 8; Page 8; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat

CC alpha-S2 casein E precursor protein.
XX
SQ Sequence 223 AA;

Query Match 83.7%; Score 77; DB 23; Length 223;
Best Local Similarity 87.5%; Pred. No. 0.00045;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHOKAMKFWIQPT 16
| | | | | | | | | | | | | | | | | |
DB 199 VDOHQKAMKFWIQPT 214

RESULT 14
AAE17471
ID AAE17471 standard; Protein; 223 AA.

AC AAE17471;

DT 22-APR-2002 (first entry)

DE Capra hircus alpha-S2 casein C precursor protein.

KW Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
KW chewing gum; cosmetic; wrinkling; periodontal disease.

OS Capra hircus.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
PT alpha-S2 casein precursor in the manufacture of a medicament for
PT alleviating or preventing periodontal disease and an effect of aging in
PT skin -

PS Claim 8; Page 8; 27pp; English.

CC The invention relates to a composition comprising a peptide or its
CC derivative. The peptide contains an amino acid sequence from alpha-S2
CC casein precursor. The peptides stimulate the growth of fibroblasts, and
CC thus the synthesis and secretion of collagen. The peptides also stimulate
CC the growth of keratinocytes, which aid in the formation and regeneration
CC of skin surface. The peptide is useful in the manufacture of a medicament
CC in the form of a toothpaste or a chewing gum, for alleviating or
CC preventing periodontal disease and a medicament in the form of a cosmetic
CC composition for alleviating or preventing an effect of aging,
CC particularly wrinkling of the skin. The present sequence is goat
CC alpha-S2 casein C precursor protein.

SQ Sequence 223 AA;

Query Match 83.7%; Score 77; DB 23; Length 223;
Best Local Similarity 87.5%; Pred. No. 0.00045;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHOKAMKFWIQPT 16
| | | | | | | | | | | | | | | | | |
DB 199 VDOHQKAMKFWIQPT 214

RESULT 15
AAE17473
ID AAE17473 standard; Protein; 223 AA.

AC AAE17473;

DT 22-APR-2002 (first entry)

DE Sheep alpha-S2 casein precursor protein.

KW Sheep; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
KW chewing gum; cosmetic; wrinkling; periodontal disease.

OS Ovis sp.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
PT alpha-S2 casein precursor in the manufacture of a medicament for
PT alleviating or preventing periodontal disease and an effect of aging in
PT skin -

PS Claim 8; Page 9; 27pp; English.

CC The invention relates to a composition comprising a peptide or its
CC derivative. The peptide contains an amino acid sequence from alpha-S2
CC casein precursor. The peptides stimulate the growth of fibroblasts, and
CC thus the synthesis and secretion of collagen. The peptides also stimulate
CC the growth of keratinocytes, which aid in the formation and regeneration
CC of skin surface. The peptide is useful in the manufacture of a medicament
CC in the form of a toothpaste or a chewing gum, for alleviating or
CC preventing periodontal disease and a medicament in the form of a cosmetic
CC composition for alleviating or preventing an effect of aging,
CC particularly wrinkling of the skin. The present sequence is sheep
CC alpha-S2 casein precursor protein.

SQ Sequence 223 AA;

Query Match 83.7%; Score 77; DB 23; Length 223;
Best Local Similarity 87.5%; Pred. No. 0.00045;
Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VYQHOKAMKFWIQPT 16
| | | | | | | | | | | | | | | | | |
DB 199 VDOHQKAMKFWIQPT 214

Search completed: July 30, 2003, 16:23:26
Job time : 33.5238 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 35.625 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-2

Perfect score: 119
Sequence: 1 VYQHQAAMKRPWIDPKTVIPY 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

sub-processing: minimum match 0%
 . Maximum Match 100%

Listing first 45 summaries

Database :

```

1: SP archaea: *
2: SP archaea: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriap: *
17: sp_archaeap: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	103	86.6	210	6	062825	062825 bubalus bub
2	95	79.8	223	6	Q9MYU6	Q9MYU6 capra hircu
3	86	72.3	223	6	Q9TQ07	Q9TQ07 capra hircu
4	86	72.3	223	6	Q9GK07	Q9GK07 capra hircu
5	52	43.7	80	10	081534	081534 saccharum o
6	52	43.7	85	10	081399	081399 orobanche c
7	52	43.7	340	10	Q96568	Q96568 ipomoea pur
8	52	43.7	379	10	Q04065	Q04065 perilla fru
9	52	43.7	386	10	Q23923	Q23923 digitalis l
10	52	43.7	388	10	Q9F637	Q9F637 torenia hyb
11	52	43.7	389	10	08LP19	08LP19 nierenbergi
12	52	43.7	389	10	Q43040	Q43040 pecunia hyb
13	52	43.7	389	10	Q93XP8	Q93XP8 nicotiana t
14	52	43.7	389	10	Q9MSB2	Q9MSB2 pecunia hyb
15	52	43.7	390	10	Q9LRB2	Q9LRB2 scutellaria
16	52	43.7	390	10	Q48564	Q48564 scutellaria

17	52	43.7	390	10	Q22122	Q12122 scutellarin
18	50	42.0	311	5	Q19257	Q19257 caenorhabdi
19	50	42.0	693	11	Q9WV66	Q9WV66 mus musculi
20	50	42.0	1877	3	Q9USK3	Q9USK3 schizosacch
21	49	41.2	282	5	Q40074	Q40074 ascaris suu
22	49	41.2	342	16	Q8WV06	Q8WV06 anabaena sp
23	48	40.3	225	4	Q9BTR9	Q9BTR9 homo sapien
24	48	40.3	328	16	Q8CE60	Q8CE60 staphylococ
25	48	40.3	374	10	Q81476	Q81476 brassica na
26	48	40.3	392	10	Q93YX5	Q93YX5 vitis sp. c
27	48	40.3	392	10	Q9PEK7	Q9PEK7 vitis labrau
28	48	40.3	392	10	Q8LEP4	Q8LEP4 parthenocis
29	48	40.3	392	10	Q9FRW2	Q9FRW2 vitis ripar
30	48	40.3	392	10	Q8BWP2	Q8BWP2 vitis ripar
31	48	40.3	392	10	Q8LEP2	Q8LEP2 parthenocis
32	48	40.3	392	10	Q9S982	Q9S982 vitis. scill
33	48	40.3	392	10	Q8LEP3	Q8LEP3 classus rhom
34	48	40.3	392	10	Q94058	Q94058 vitis vinif
35	48	40.3	392	10	Q944W8	Q944W8 vitis sp. c
36	48	40.3	392	10	Q944W7	Q944W7 vitis sp. c
37	48	40.3	394	10	Q93V86	Q93V86 humulus lup
38	48	40.3	395	10	Q94FW0	Q94FW0 alliarria pe
39	48	40.3	399	10	Q94FW8	Q94FW8 humulus lup
40	48	40.3	408	16	Q9SLY0	Q9SLY0 psilocum nu
41	48	40.3	418	16	Q9WTV5	Q9WTV5 thermocoga
42	48	40.3	704	4	Q9H992	Q9H992 homo sapien
43	47	39.5	56	10	Q8R250	Q8R250 vaccinium m
44	47	39.5	80	10	Q9S856	Q9S856 glycine max
45	47	39.5	115	10	Q9AXX8	Q9AXX8 brassica ol

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	210 AA.
ID 062825			
AC 062825;			
DT 01-AUG-1998 (TREMBLrel. 07, Created)			
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE AS2-casein (Fragment).			
OS Bubalus bubalis (Domestic water buffalo).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC Bovidae; Bovinae; Bubalus.			
OX NCBI_TaxID=89462;			
XX [1]			
RN SEQUENCE FROM N.A.			
RP TISSUE=Mammary gland;			
RA Das P., Jain S., Gary L.C.;			
RT "Cloning and nucleotide sequence of cDNA encoding as2-casein in B.			
RT bubalis".			
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.			
DR EMBL; AJ005431; CAA06534.2; -.			
DR InterPro; IPR001588; Casein.			
DR Pfam; PF00363; caseins; 2.			
DR NON TER			
FT 1			
SO SEQUENCE	210 AA;	24700 MM;	05DEF95963F1132C CXC64;
Query Match	86.6%;	Score 103;	DB 6;
Best Local Similarity	85.7%;	Pred. No. 5.1e-08;	Length 210;
Matches 18;	Conservative 1;	Mismatches 2;	Indels 0;
Gaps 0;			
Qy 1 VYGHQAKMKPPIQPKTKVIPP 21			
Db 186 VYGYOKAMKPPWQPKTNVIPP 206			
RESULT 2			
Q9MYU6	PRELIMINARY;	PRT;	223 AA.
Q9MYU6;			

DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Alpha s2-casein.
 GN CSN152.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OC NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lagomigro R., Pilla F., Matassino D., Zullo A.;
 RT "A new allele of goat alpha s2-casein gene."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ289716; CAB94236.1; -
 DR EMBL; AJ297313; CAC21704.2; JOINED.
 DR InterPro; IPR001588; Casein.
 DR Pfam; PF00363; caseins; 2.
 DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 SQ SEQUENCE 223 AA; 26403 MW; 0E1FE83F24DA85E2 CRC64;

Query Match 79.8%; Score 95; DB 6; Length 223;
 Best Local Similarity 81.0%; Pred. No. 9.3e-07;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VYOHKAMKRWIOPTKVIPIY 21
 Db 199 VDOHOKAMKRWIOPTKNAIPY 219

RESULT 3

ID 09T07 PRELIMINARY; PRT; 223 AA.
 AC 09T07;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Alpha s2-casein.
 GN CSN152.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OC NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veltri C.C., Pilla F., Lagomigro R.R.;
 RT "A new allele of goat alpha s2-casein."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ249995; CAB59920.1; -
 DR InterPro; IPR001588; Casein.
 DR Pfam; PF00363; caseins; 2.
 DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 SQ SEQUENCE 223 AA; 26433 MW; CE9FADCD7688293 CRC64;

Query Match 72.3%; Score 86; DB 6; Length 223;
 Best Local Similarity 76.2%; Pred. No. 2.3e-05;
 Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VYOHKAMKRWIOPTKVIPIY 21
 Db 199 VDOHOKAMKRWIOPTKNAIPY 219

RESULT 4

ID 09G07 PRELIMINARY; PRT; 223 AA.
 AC 09G07;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Alpha s2-casein.
 GN CSN152.

OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OC NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veltri C., Pilla F., Lagomigro R.;
 RT "A new allele of alpha s2-casein."
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ297310; CAC21704.2; -
 DR EMBL; AJ297311; CAC21704.2; JOINED.
 DR EMBL; AJ242728; CAC21704.2; JOINED.
 DR EMBL; AJ297312; CAC21704.2; JOINED.
 DR EMBL; AJ297313; CAC21704.2; JOINED.
 DR EMBL; AJ297314; CAC21704.2; JOINED.
 DR EMBL; AJ242527; CAC21704.2; JOINED.
 DR EMBL; AJ297315; CAC21704.2; JOINED.
 DR EMBL; AJ297316; CAC21704.2; JOINED.
 DR EMBL; AJ242526; CAC21704.2; JOINED.
 DR EMBL; AJ242528; CAC21704.2; JOINED.
 DR EMBL; AJ242533; CAC21704.2; JOINED.
 DR InterPro; IPR001588; Casein.
 DR Pfam; PF00363; caseins; 2.
 DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 SQ SEQUENCE 223 AA; 26432 MW; CE97658D7688C9D CRC64;

Query Match 72.3%; Score 86; DB 6; Length 223;
 Best Local Similarity 76.2%; Pred. No. 2.3e-05;
 Matches 16; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VYOHKAMKRWIOPTKVIPIY 21
 Db 199 VDOHOKAMKRWIOPTKNAIPY 219

RESULT 5

ID 081534 PRELIMINARY; PRT; 80 AA.
 AC 081534;
 DT 01-NOV-1998 (TREMBlrel. 08, Created)
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Chalcone synthase (Fragment).
 GN CHS.
 OS Saccharum officinarum (Sugarcan).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
 OC NCBI_TaxID=4547;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veronesi C., Thaliouarn P.;
 RT "Sugarcan resistance to Striga hermonthica."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF079174; AAC27797.1; -
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_scl1_synth; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 FT NON TER 1 1
 FT NON TER 80 80
 SQ SEQUENCE 80 AA; 8744 MW; 7B94776D2A46F418 CRC64;

Query Match 43.7%; Score 52; DB 10; Length 80;
 Best Local Similarity 64.3%; Pred. No. 1.4;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 5 OKAMKRWIOPTKVIPIY 18
 Db 33 OKAIRKRWIOPTKVIPIY 46

RESULT 6			
ID	081399	PRELIMINARY;	PRT; 85 AA.
AC	081399;		
DT	01-NOV-1998 (TrEMBLrel. 08, Created)		
DT	01-NOV-1998 (TrEMBLrel. 08, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Chalcone synthase (Fragment).		
GN	CHS.		
OS	Orobanchaceae cumana.		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;		
OC	Ascertidae; Lamiales; Orobanchaceae; Orobanchaceae; Orobanchaceae		
OX	NCBI_TaxID=78542;		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RT	Veroneste C., Labrousse P., Thallouarn P.;		
RT	"Histological, physiological and molecular aspects of sunflower		
RT	resistance to Orobanchaceae cumana.";		
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF074401; AAC26132.1; -.		
DR	HSSP; P30074; 1D6F		
DR	InterPro; IPR001099; N-C synthase.		
DR	Pfam; PF00195; Chal_still_synth; 1.		
DR	ProDom; PD000453; N-C synthase; 1.		
FT	NON_TER	1	85
FT	NON_TER	85	85
SQ	SEQUENCE	85 AA; 9483 MW; B9A6ABEDD6D4777E CRC64;	

Query Match	43.7%	Score 52;	DB 10;	length 85;
Best Local Similarity	64.3%	Pred. No. 1.5;		
Matches	9;	Conservative	3;	Mismatches 2;
				Indels 0;
				Gaps 0

```
QY      5 QKAMKPMIOPKTKV 18
         |||:| |||:|:|
Db     33 QKAIKEWGQPKSKI 46
```

```

RESULT 7
096568
ID 096568 PRELIMINARY; PRT; 340 AA.
AC 096568;
DT 01-FEB-1997 (T-EMBLrel. 02, Created)
DT 01-FEB-1997 (T-EMBLrel. 02, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Chalcone synthase.
NN CHS-FL1.

```

OC Eukaryota; Eukaryota; "eukaryotic group";
OC Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Lamiales; Solanales; Convolvulaceae; Ipomoea.
OX NCBI_TaxID=4121;
OX (1)
RP SEQUENCE FROM N.A.
RP TISSUE=Flower buds;
RC Rauscher M.D., Tiffin P.L., Miller R.E.;
RA "Regulation of anthocyanin gene expression in *Ipomoea purpurea*.";
RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RL

DR HSSP, P30074; IDGF, N-C-synthase.
DR InterPro, IPR001059; N-C-synthase.
DR Pfam, PF00195; Cal_s11_synt, 1.
DR Pfam, PF02797; Cal_s11_syntc, 1.
DR ProDom, PD000453; N-C-synthase, 1.
DR ProSite, PS00041; CALICONE_SYNTN, 1.
DR Q03931; 340 AA; 37519 MW; 0668266873DCEA93 CRC64
Q0 SEQUENCE

Query Match	43.7%	Score 52;	DB 10;	Length 340;
Best Local Similarity	64.3%	Pred. No. 6.2;		
Matches	9;	Conservative	2;	Indels 0; Gaps 0;

5 QKAMKPWIOPTKV 18
|||:| | |||:|:
QY

Db 111 QKAIKEWGQPKSKI 124

RESULT 8

DT 01-JUL-1997 (T-EMBLrel. 04, Created)
DT 01-JUL-1997 (T-EMBLrel. 04, last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, last annotation update)
DE Chalcone synthase.
OS *Penttilia frutescens*.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
OC Asteridae; lamiales; Lamiales; lamiaceae; Nepetoideae; Elaeagnaceae
OC *Penttilia*

OX NCBI_TaxID=48386;
RN
RP
SEQUENCE FROM N.A.
RX MEDLINE=98088016; PubMed=9426610;
RAY G. Z., Yamazaki M., Sugiyama M., Tanaka Y., Saito K.;
RT "Cloning and molecular analysis of structural genes involved in
anthocyanin biosynthesis and expressed in a forma-specific manner in
RT *Perilla frutescens*."
RL Plant Mol. Biol. 35:915-927(1997).
DR EMBL; AF002582; BAI19546.1; -.
DR HSSP; P30074; 1DEP.
DR InterPro; IPR001099; N-C_synthase.
DR Pfam; PF00195; Chal_s11_synthc; 1.
DR Pfam; PF02797; Chal_s11_synthc; 1.
DR Prodom; PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
SQ SEQUENCE 379 AA; 41380 MW; SE149C5530BD106 CRC64;

Query Match	43.7%	Score 52:	DB 10;	length 379;
Best Local Similarity	64.3%	Pred. NC.	6.9;	
Matches 9;	Conservative 3;	Mismatches 2;	Indels 0;	Gaps 0

```
QY      5 QKAMKPWIQPKTKV 18
         |||:| | |||:|
Db      111 QKAIKEWGPCKSKI 124
```

RESULT 9
023923
ID 023923 PRELIMINARY; PRT; 386 AA

DT 01-JAN-1998 (T-EMBLrel. 05, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Chalcone synthase (Fragment).
OS Digitalis lanata (Foxglove).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Lamiales; Antirrhinaceae; Digitalis
NCBI_TaxID=949450;
OX

RP SEQUENCE FROM N.A.
RC TISSUE=leaf;
RA Thoenigster C., Lindemann P., Iuckner M.;
RT "Expression of phenylalanine ammonia lyase, chalcone synthase and
RT phytochrome in somatic embryogenesis of *Digitalis lanata*";
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ
RL EMBL, AJ002526; CAA0512.1; --

DR InterfPro019p0001099; N-C-synthase.
DR pfem1; P001156; Chal_sgl1_synC; 1.
DR pfem1; P001797; Chal_sgl1_synC; 1.
DR ProDom; P0000453; N-C-synthase; 1.
DR PROSITE; PS000441; CHALCONO_SYNTH; 1.
NON_TER 1
SEQ 386 AA; 42190 MW; 0A8EBBF6EDBA82F8F CRC64

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 ; Search time 6.875 Seconds
(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-2

Perfect score: 119
Sequence: 1 VYQHOKAMKPMIOPKTVIPY 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	222	1	CAS2_BOVIN
2	95	79.8	223	1	CAS2_CAPII
3	95	79.8	223	1	CAS2_SHEEP
4	59.5	50.0	235	1	CAS2_PIG
5	53	49.6	182	1	CAS3_RABIT
6	52	43.7	193	1	CAS2_CAMDR
7	52	43.7	389	1	CHS1_LYCES
8	52	43.7	389	1	CHS2_LYCES
9	52	43.7	389	1	CHS2_SOUTU
10	52	43.7	389	1	CHSA_PETHY
11	52	43.7	389	1	CHSA_SOUTU
12	52	43.7	389	1	CHSB_SOUTU
13	52	43.7	389	1	CHSE_TPOPU
14	52	43.7	389	1	CHSE_PHANI
15	52	43.7	390	1	CHSY_PETHY
16	52	43.7	390	1	CHSY_ANTMA
17	52	43.7	391	1	CHSY_PPRFR
18	52	43.7	400	1	CHS7_SORBI
19	52	43.7	401	1	CHS1_SORBI
20	52	43.7	401	1	CHS2_SORBI
21	52	43.7	401	1	CHS3_SORBI
22	52	43.7	401	1	CHS4_SORBI
23	52	43.7	401	1	CHS6_SORBI
24	51	42.9	180	1	CAS2_RABIT
25	50	42.0	291	1	PCIT_MOUSE
26	49	41.2	207	1	VEGB_MOUSE
27	49	41.2	392	1	CHS1_SEBCE
28	49	41.2	394	1	CHS2_SEBCE
29	49	41.2	400	1	CHS1_HORVU
30	49	41.2	400	1	CHS2_MAIZE
31	48	40.3	389	1	THS3_VITVI
32	48	40.3	392	1	THS1_VITVI
33	48	40.3	392	1	THS2_VITVI

ALIGNMENTS

RESULT 1	ID	CAS2_BOVIN	STANDARD;	PRT;	222 AA.
AC	P02663	Q9TR51			
DT	21-JUL-1986	(Rel. 01, Created)			
DT	01-UTL-1989	(Rel. 11, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Alpha-S2 casein precursor [Contains: Casocidin-1].				
GN	CSN1S2.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID=9913;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=88188989; PubMed=2833669;				
RA	Stewart A.F., Bomsing J., Beattie C.W., Shah F., Wallis I.M.,				
RA	Mackinlay A.G.;				
RT	"Complete nucleotide sequences of bovine alpha S2- and beta-casein				
RT	CDNA: comparisons with related sequences in other species.";				
RL	Mol. Biol. Evol. 4:231-241(1987).				
RN	[2]				
RP	SEQUENCE OF 16-222 (A ALLELE).				
RC	TISSUE=Milk;				
RX	MEDLINE=77185633; PubMed=862906;				
RA	Birgion G., Ribadeau-Dumas B., Mercier J.-C., Pellesier J.-P.,				
RA	Das B.C.;				
RT	"Complete amino acid sequence of bovine alphaS2-casein.";				
RL	FEBS Lett. 76:274-279(1977).				
RN	[3]				
RP	PARTIAL SEQUENCE (D ALLELE).				
RX	TISSUE=Milk;				
RC	MEDLINE=79239837; PubMed=469044;				
RA	Groesclaude F., Joudrier P., Mahe M.-F.;				
RT	"A genetic and biochemical analysis of a polymorphism of bovine alpha				
RT	S2-casein.";				
RL	J. Dairy Res. 46:211-213(1979).				
RN	[4]				
RP	SEQUENCE OF 165-203, AND CHARACTERIZATION OF CASOCIDIN.				
RC	TISSUE=Milk;				
RX	MEDLINE=96000204; PubMed=7556666;				
RA	Zucht H.-D., Ralda M., Adermann K., Meagert H.-U., Forstmann W.-G.;				
RT	"Casocidin-1: a casein-alpha s2 derived peptide exhibits antibacterial				
RT	activity.";				
RL	FEBS Lett. 372:185-188(1995).				
CC	-1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT				
CC	CALCIUM PHOSPHATE.				
CC	-1- FUNCTION: CASOCIDIN-1 INHIBITS THE GROWTH OF ESCHERICHIA COLI AND				
CC	STAPHYLOCOCCUS CARNOUS.				
CC	-1- SUBCELLULAR LOCATION: Extracellular.				
CC	-1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.				
CC	-1- MASS SPECTROMETRY: MW=4870; METHOD=Electrospray; RANGE=165-203.				
CC	-1- POLYMORPHISM: AT LEAST TWO ALLELES EXIST. THE SEQUENCE OF THE A				
CC	ALLELE IS SHOWN HERE. THE D ALLELE SEQUENCE DIFFERS FROM THAT				
CC	SHOWN IN HAVING A DELETION OF NINE RESIDUES, WHICH MAY BE 49-58,				

34	47	39.5	285	1	CHS6_MEDSA	P51079 medicago sa
35	47	39.5	311	1	CHS4_TRISTU	P51086 trifolium s
36	47	39.5	388	1	CHS1_SOYBN	P24826 glycine max
37	47	39.5	388	1	CHS2_SOYBN	P17957 glycine max
38	47	39.5	388	1	CHS3_SOYBN	P19168 glycine max
39	47	39.5	388	1	CHS5_SOYBN	P48406 glycine max
40	47	39.5	388	1	CHS6_SOYBN	P30080 glycine max
41	47	39.5	388	1	CHS8_IPOBA	Q9mb36 ipomoea bat
42	47	39.5	388	1	CHSD_PHANI	O22045 pharbitis n
43	47	39.5	389	1	CHS_VIGDN	P51089 vigna ungui
44	47	39.5	389	1	CHS1_CAMSI	P48386 camellia si
45	47	39.5	389	1	CHS1_CASGL	Q9zrr8 casuarina g

CC 50-59, OR 51-60.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -1- DATABASE: NAME-Protein Spotlight;
 CC NOTE: Issue 16 of November 2001;
 CC WWW: <http://www.expasy.org/spotlight/articles/sp1c016.html>.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M16644; AAA30479.1; -.
 CC PIR: J02008; KABOS2.
 CC DR InterPro: IPR001588; Casein.
 CC DR Pfam: PF00363; caseins; 2.
 CC DR PROSITE: PS00306; CASEIN ALPHA BETA; 1.
 CC KM Milk; Phosphorylation; Signal; Repeat; Antibiolic.
 CC FT SIGNAL 1 15
 CC FT CHAIN 16 222 ALPHA-S2 CASEIN.
 CC FT PEPTIDE 165 203 CASOCIDIN-I.
 CC FT REPEAT 76 140
 CC FT MOD_RES 23 23 PHOSPHORYLATION.
 CC FT MOD_RES 24 24 PHOSPHORYLATION.
 CC FT MOD_RES 25 25 PHOSPHORYLATION.
 CC FT MOD_RES 31 31 PHOSPHORYLATION.
 CC FT MOD_RES 71 71 PHOSPHORYLATION.
 CC FT MOD_RES 72 72 PHOSPHORYLATION.
 CC FT MOD_RES 73 73 PHOSPHORYLATION.
 CC FT MOD_RES 76 76 PHOSPHORYLATION.
 CC FT MOD_RES 144 144 PHOSPHORYLATION.
 CC FT MOD_RES 146 146 PHOSPHORYLATION.
 CC FT CONFLICT 102 102 Q -> E (IN REF. 2).
 CC SQ SEQUENCE 222 AA; 26019 MW; 81E7408AF1C12F7C CRC64;
 CC
 CC Query Match 100.0%; Score 119; DB 1; Length 222;
 CC Best Local Similarity 100.0%; Pred. No. 9, 2e-11;
 CC Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 1 VYOHQKAMKPMWIOPTKVIPIY 21
 CC Db 198 VYOHQKAMKPMWIOPTKVIPIY 218
 CC
 CC RESULT 2
 CC CAS2_CAPHI STANDARD; PRT; 223 AA.
 CC ID CAS2_CAPHI
 CC AC P33049;
 CC DT 01-OCT-1993 (Rel. 27, Created)
 CC DT 01-OCT-1993 (Rel. 27, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Alpha-S2 casein precursor (Alpha-S2-CN).
 CC GN CSN182.
 CC OS Capra hircus (Goat).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC OC Bovidae; Caprinae; Capra.
 CC OC NCB1_TaxID=9925;
 CC OX [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=93216130; PubMed=8462880;
 CC RA Bouniol C.;
 CC RT "Sequence of the goat alpha s2-casein-encoding cDNA."
 CC RL Gene 125:235-236 (1993).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RA Bouniol C., Brignon G., Mahe M.-F., Printz C.;
 CC RT "Characterization of goat allelic alpha-s2-caseins A and B: further
 CC evidence of the phosphorylation code of caseins";
 CC RL Protein Seq. Data Anal. 5:213-218 (1993).
 CC

RN [3]
 RP SEQUENCE FROM N.A. (VARIANT C).
 RX MEDLINE=95030556; PubMed=7943951;
 RA Bouniol C., Brignon G., Mahe M.-F., Printz C.;
 RT "Biochemical and genetic analysis of variant C of caprine alpha s2-
 RT casein (Capra hircus)."
 RL Anim. Genet. 25:173-177 (1994).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC CALCIUM PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- POLYMORPHISM: THREE ALLELES OF ALPHA-S2 CASEIN ARE KNOWN: A, B AND
 CC C. THE FREQUENCIES OF THE ALLELES IS ESTIMATED TO BE 0.85, 0.04
 CC AND 0.11 IN THE FRENCH DAIRY BREEDS 'ALPINE' AND 'SAANEN'.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X65160; CAA46278.1; -.
 CC PIR: S74171; AAB32166.1; -.
 CC DR PIR: I46995; I46995.
 CC DR PIR: S33881; JN0547.
 CC DR InterPro: IPR001588; Casein.
 CC DR Pfam: PF00363; caseins; 2.
 CC DR PROSITE: PS00306; CASEIN ALPHA BETA; 1.
 CC KM Milk; Phosphorylation; Signal; Repeat; Polymorphism.
 CC FT SIGNAL 1 15
 CC FT CHAIN 16 223 ALPHA-S2 CASEIN.
 CC FT REPEAT 77 141
 CC FT MOD_RES 23 23
 CC FT MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 77 77 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 145 145 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 147 147 PHOSPHORYLATION (BY SIMILARITY).
 CC FT VARIANT 79 79 E -> K (IN VARIANT B).
 CC FT VARIANT 182 182 K -> I (IN VARIANT C).
 CC SQ SEQUENCE 223 AA; 26389 MW; 187DE42FD688291 CRC64;
 CC
 CC Query Match 79.8%; Score 95; DB 1; Length 223;
 CC Best Local Similarity 81.0%; Pred. No. 3, 6e-07;
 CC Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 CC
 CC QY 1 VYOHQKAMKPMWIOPTKVIPIY 21
 CC Db 199 VYOHQKAMKPMWIOPTKVIPIY 219
 CC
 CC RESULT 3
 CC CAS2_SHEEP STANDARD; PRT; 223 AA.
 CC ID CAS2_SHEEP
 CC AC P04654;
 CC DT 13-AUG-1987 (Rel. 05, Created)
 CC DT 13-AUG-1987 (Rel. 05, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Alpha-S2 casein precursor.
 CC GN CSN182.
 CC OS Ovis aries (Sheep).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC OC Bovidae; Caprinae; Ovis.
 CC OC NCB1_TaxID=9940;
 CC OX [1]
 CC

RP SEQUENCE FROM N.A.
RA MEDLINE=86104467; PubMed=3002499;
RX Boissard M., Petrisant G.;
RT "Complete sequence of ovine alpha s2-casein messenger RNA.";
RL Biochimie 67:1043-1051(1985).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X03238; CAA26983.1; -.
CC PIR; A25070; KASHS2.
CC InterPro: IPR001588; Casein.
CC Pfam: PF00363; caseins; 2.
CC PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
CC KW Milk; Phosphorylation; Repeat; Signal.
CC FT SIGNAL 1 15
CC FT CHAIN 16 223 ALPHA-S2 CASEIN.
CC FT REPEAT 77 141
CC FT REPEAT 159 223
CC FT MOD_RES 23 23 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 24 24 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 25 25 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 72 72 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 73 73 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 74 74 PHOSPHORYLATION (POTENTIAL).
CC FT VARIANT 64 64 D -> N.
CC SQ SEQUENCE 223 AA; 26332 MW; 67212935E27426D7 CRC64;
CC -----
CC Query Match 79.8%; Score 95; DB 1; Length 223;
CC Best Local Similarity 81.0%; Pred. No. 3.6e-07;
CC Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
CC -----
CC QY 1 VYQHOKAMKPMWIOPTKVIPIY 21
CC DB 199 VQHOKAMKPMWIOPTKVIPIY 219
CC -----
CC RESULT 4
CC CAS2 PIG STANDARD; PRT; 235 AA.
CC ID CAS2 PIG
CC AC P39036;
CC DT 01-FEB-1995 (Rel. 31, Created)
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Alpha-S2 casein precursor.
CC GN CSN1S2.
CC OS Sus scrofa (Pig).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sub.
CC OX NCBI_TaxID=9823;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC TISSUE=Mammary gland;
CC RA MEDLINE=92367960; PubMed=1503276;
CC RA Alexander L.J., Das Gupta N.A., Beattie C.W.;
CC RT "The sequence of porcine alpha s2-casein cDNA.";
CC RT Anim. Genet. 23:365-367(1992).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54975; CAA38719.1; -.
CC InterPro: IPR001588; Casein.
CC Pfam: PF00363; caseins; 2.
CC PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
CC KW Milk; Phosphorylation; Signal.
CC FT SIGNAL 1 15
CC FT CHAIN 16 235 ALPHA-S2 CASEIN.
CC FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 70 70 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 71 71 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
CC SQ SEQUENCE 235 AA; 27570 MW; C903B760D184C14C CRC64;
CC -----
CC Query Match 50.0%; Score 59.5; DB 1; Length 235;
CC Best Local Similarity 56.5%; Pred. No. 0.08;
CC Matches 13; Conservative 3; Mismatches 4; Indels 3; Gaps 1;
CC -----
CC QY 1 VYQHOKAMKPMWIOPTKVIPIY 20
CC DB 208 VQHOKAMKPMWIOPTKVIPIY 230
CC -----
CC RESULT 5
CC CAS3 RABIT STANDARD; PRT; 182 AA.
CC ID CAS3 RABIT
CC AC P50419;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 01-OCT-1996 (Rel. 34, Last annotation update)
CC DE Alpha-S2B casein precursor.
CC OS Oryctolagus cuniculus (Rabbit).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
CC OX NCBI_TaxID=9986;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=New Zealand white; TISSUE=Mammary gland;
CC RX MEDLINE=94107245; PubMed=8280077;
CC RA Dawson S.P., Wilde C.J., Tighe P.O., Mayer R.J.;
CC RT "Characterization of two novel casein transcripts in rabbit mammary
CC gland.";
CC RL Biochem. J. 296:777-784(1993).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X76909; CAA54231.1; -.
CC PIR; S39776; S39776.
CC InterPro: IPR001588; Casein.
CC Pfam: PF00363; caseins; 2.
CC -----

DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 KW Milk; Phosphorylation; Signal.
 FT SIGNAL
 FT CHAIN 16 182 ALPHA-S2B CASEIN.
 SQ SEQUENCE 182 AA; 21597 MW; F2433C2DA45550C CRC64;
 Query Match 49.6%; Score 59; DB 1; Length 182;
 Best Local Similarity 42.9%; Pred. No. 0.074;
 Matches 9; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 VYOHQKMKPMIOPTKVIPY 21
 DB 83 LYQIPYVMDPMTAEKTAIP 103

RESULT 6
 CAS2_CAMDR STANDARD; PRT; 193 AA.
 AC 097944;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Alpha-S2 casein precursor.
 GN CSN152.
 OS Camelus dromedarius (Dromedary) (Arabian camel).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 ON NCBI_TaxID=9838;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN=Somali; TISSUE=Udder;
 RX MEDLINE=98291310; PubMed=5627840;
 RA Kappeler S., Farah Z., Puhus Z.;
 RT "Sequence analysis of Camelus dromedarius milk caseins";
 RJ J. Dairy Res. 65:209-222(1998).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC CALCIUM PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR 1 EMBL; AJ012629; CA10078.1; -
 DR InterPro; IPR001588; Casein.
 DR Pfam; PF00363; caseins; 2.
 DR PROSITE; PS00306; CASEIN_ALPHA_BETA; FALSE_NEG.
 KW Milk; Phosphorylation; Signal.
 FT SIGNAL 1 15
 FT CHAIN 16 193 ALPHA-S2 CASEIN.
 SQ SEQUENCE 193 AA; 22964 MW; 2843256F8FD2ED13 CRC64;
 Query Match 43.7%; Score 52; DB 1; Length 193;
 Best Local Similarity 42.9%; Pred. No. 0.87;
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 VYOHQKMKPMIOPTKVIPY 21
 DB 93 LHQCQVIMNPFWDGKTAPYEP 113

RESULT 7
 CHS1_LYCES STANDARD; PRT; 389 AA.
 AC P23418;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1).
 GN CHS1.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Solanum.
 ON NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cotyledon, Hypocotyl, and Leaf;
 RX MEDLINE=91117196; PubMed=1960524;
 RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
 RT "Molecular genetic analysis of chalcone synthase in Lycopersicon
 RT esculentum and an anthocyanin-deficient mutant."
 RL Mol. Gen. Genet. 224:279-288(1990).
 CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X55194; CA38980.1; -
 DR InterPro; IPR001099; N-C-synthase.
 DR Pfam; PF00195; Chal_gtl_synth; 1.
 DR Pfam; PF02797; Chal_gtl_synth; 1.
 DR PRODOM; PD000453; N-C synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 KW Flavonoid biosynthesis; Transferase; Acyltransferase;
 KW Multigene family.
 FT ACT SITE 164 164 BY SIMILARITY.
 FT SITE 164 164 BY SIMILARITY.
 SQ SEQUENCE 389 AA; 42552 MW; 553DC695EBA96A8B CRC64;
 Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
 DB 111 QXAKEMGPKSKI 124

RESULT 8
 CHS2_LYCES STANDARD; PRT; 389 AA.
 AC P23419;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
 GN CHS2.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Solanaceae; Solanum.
 ON NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cotyledon, Hypocotyl, and Leaf;
 RX MEDLINE=91117196; PubMed=1960524;

RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
 RT "Molecular genetic analysis of chalcone synthase in *Lycopersicon*
 RL *esculentum* and an anthocyanin-deficient mutant.";
 CC Mol. Gen. Genet. 224:279-288(1990).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; X55195; CAA38981.1; -.
 CC DR InterPro; IPR001099; N-C_synthase.
 CC DR Pfam; PF00195; Chal_act1_synth; 1.
 CC DR Pfam; PF02797; Chal_act1_synth; 1.
 CC DR Prodom; PD000453; N-C_synthase; 1.
 CC DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 CC KW Flavonoid biosynthesis; Transferase; Acyltransferase;
 CC KM Multigene family.
 CC FT ACT SITE 164
 CC SQ SEQUENCE 389 AA; 42730 MW; F92E46B3B5FC32F CRC64;
 CC -----
 CC Query Match 43.7%; Score 52; DB 1; Length 389;
 CC Best Local Similarity 64.3%; Pred. No. 1.8;
 CC Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 5 OKAMKPMIOPTKV 18
 CC |||||
 CC Db 111 OKAIKMGQPKSKI 124
 CC
 CC RESULT 9
 CC CHS2_SOLITU STANDARD; PRT; 389 AA.
 CC ID CHS2_SOLITU
 CC AC Q43188;
 CC DT 15-JUL-1999 (Rel. 38, Created)
 CC DT 15-JUL-1999 (Rel. 38, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
 CC GN CHS2.
 CC OS Solanum tuberosum (Potato).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 CC OX NCBI_Taxid=4113;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX STRAIN=cv. Red Pontiac;
 CC MEDLINE=97141614; PubMed=8987872;
 CC RA Jeon J.-H., Kim H.S., Choi K.H., Jeong Y.H., Byun S.-M.;
 CC RT "Cloning and characterization of one member of the chalcone synthase
 CC RL gene family from *Solanum tuberosum* L.";
 CC CC Biosci. Biotechnol. Biochem. 60:1907-1910(1996).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY

CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; U47738; AAB05239.1; -.
 CC DR PIR; JCS136; JCS136.
 CC DR InterPro; IPR001099; N-C_synthase.
 CC DR Pfam; PF00195; Chal_act1_synth; 1.
 CC DR Pfam; PF02797; Chal_act1_synth; 1.
 CC DR Prodom; PD000453; N-C_synthase; 1.
 CC DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 CC KW Flavonoid biosynthesis; Transferase; Acyltransferase;
 CC KM Multigene family.
 CC FT ACT SITE 164
 CC SQ SEQUENCE 389 AA; 42476 MW; 41618F944958603 CRC64;
 CC -----
 CC Query Match 43.7%; Score 52; DB 1; Length 389;
 CC Best Local Similarity 64.3%; Pred. No. 1.8;
 CC Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC QY 5 OKAMKPMIOPTKV 18
 CC |||||
 CC Db 111 OKAIKMGQPKSKI 124
 CC
 CC RESULT 10
 CC CHS4_PETHY STANDARD; PRT; 389 AA.
 CC ID CHS4_PETHY
 CC AC P08894;
 CC DT 01-NOV-1988 (Rel. 09, Created)
 CC DT 01-NOV-1988 (Rel. 09, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Chalcone synthase A (EC 2.3.1.74) (Naringenin-chalcone synthase A).
 CC GN CHS4.
 CC OS Petunia hybrida (Petunia).
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 CC OX NCBI_Taxid=4102;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX STRAIN=cv. Violet 30;
 CC MEDLINE=86286540; PubMed=3016642;
 CC RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Velkamp E.,
 CC RA Mol J.N.M.;
 CC RT "Floral tissue of *Petunia hybrida* (V30) expresses only one member of
 CC RL the chalcone synthase multigene family.";
 CC Nucleic Acids Res. 14:5229-5239(1986).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RX STRAIN=cv. Violet 30; TISSUE=Leaf;
 CC MEDLINE=90034197; PubMed=2806915;
 CC RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;
 CC RT "Cloning and molecular characterization of the chalcone synthase
 CC RL multigene family of *Petunia hybrida*.";
 CC Gene 81:245-257(1989).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- TISSUE SPECIFICITY: MAJOR EXPRESSED MEMBER OF THE GENE FAMILY IN

```

CC VARIOUS FLORAL TISSUES AND IN SEEDLINGS TREATED WITH UV LIGHT. IT
CC IS RELATIVELY LOW EXPRESSED IN TISSUE CULTURE MATERIAL.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.1bd-sib.ch/announce/
CC or send an email to license@1bd-sib.ch).
CC -----
DR EMBL, X04080; CAA27718.1; -.
DR EMBL, X14591; CAA32731.1; -.
DR PIR, A23643; SYRUCN.
DR PIR, J50308; SYPUCA.
DR InterPro, IPR001099; N-C synthase.
DR Pfam, PF00195; Chal_stil_synthc.1.
DR Pfam, PF02197; Chal_stil_synthc.1.
DR Prodom, PD000453; N-C synthase.1.
DR PROSITE, PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferrase; Acyltransferase;
KW Multigene family.
FT ACT_SITE 164 BY SIMILARITY.
FT CONFID 128 V -> F (IN REF. 2).
SQ SEQUENCE 389 AA, 42525 MW, 0221BBD41E96AFA6 CRC64;

Query Match 43.7%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 1.8;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMOPKTKV 18
Db 111 QRAIKWQGPXSKI 124
|||:|||||:|
|||:|||||:|

RESULT 11
CHSA_SOLTU STANDARD; PRT; 389 AA.
ID CHSA_SOLTU
AC Q41436;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase 1A (EC 2.3.1.74) (Naringenin-chalcone synthase 1A).
GN CHSA1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae, lamids; Solanales; Solanaceae; Solanum.
OC NCBI_TaxId=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Red Pontiac;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Young Y.-H., Young H., Byun S.-M.;
RT "Characterization of two members of the chalcone synthase gene family
RT from Solanum tuberosum L.";
RL (In) Plant Gene Register PCR96-027.
CC -1- FUNCTION: The primary product of this enzyme is 4',2'',4'',6''-
CC tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U47739; AAB67734.1; -.
DR InterPro; IPR001099; N-C_synthase.
DR Pfam; PF00195; Chal_stil_syntc; 1.
DR DR P002797; Chal_stil_syntc; 1.
DR ProDom; PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KW Multigene family.
KW ACT SITE
SO SEQUENCE 369 AA; 42562 MW; B18ID9C6B9170A14 CRC64;
OY 5 OKANKPMIOPKTV 18
Db 111 OKALKENGOPKSKT 124

RESULT 12
CHSB_SOLUTU STANDARD; PRT; 369 AA.
AC Q4163;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase 1B (EC 2.3.1.74) (Naringenin-chalcone synthase 1B).
GN CHS1B.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxId=4113;
XP [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Red Pontiac;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Young Y.-H., Young S.-M.;
RT "Characterization of two members of the chalcone synthase gene family
RT from Solanum tuberosum L. ";
RL (1n) Plant Gene Register PGR96-027.
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U47740; AAB67735.1; -.
DR InterPro; IPR001099; N-C_synthase.
DR Pfam; PF00195; Chal_stil_syntc; 1.
DR DR P002797; Chal_stil_syntc; 1.
DR ProDom; PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KW Multigene family.
KW ACT_SITE
SO SEQUENCE 369 AA; 42548 MW; E7334A1A34BDD1CC CRC64;
BY SIMILARITY
SEQUENCE 369 AA; 42548 MW; E7334A1A34BDD1CC CRC64;

```


Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAKMPWIOPKTV 18
 |||:|||||:
 DB 111 OKAIKEMGQPKSKI 124

RESULT 13

CHSE_IPOPU STANDARD; PRT; 389 AA.
 AC 022047;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase E (EC 2.3.1.74) (Naringenin-chalcone synthase E)
 DE (CHS-E).
 GN CHSE.

OS Ipomoea purpurea (Common morning glory).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Convolvulaceae; Ipomoea.

OC NCBI_TaxID=4121;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. FR-35; Tissue=Flower buds;
 RX MEDLINE=97393496; PubMed=9249990;

RA Fukuda-Tanaka S., Hoshino A., Hisatomi Y., Habu Y., Hasebe M.,
 RA Iida S.;

RT Identification of new chalcone synthase genes for flower pigmentation
 in the Japanese and common morning glories.";

RL Plant Cell Physiol. 38:754-758(1997).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
 which can under specific conditions spontaneously isomerize into
 naringenin.

CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 OF WHICH ARE BRIGHTLY COLORED.

CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AB001827; BAA21789.1; -
 CC PIR; T07799; T07799.

DR InterPro; IPR001099; N-C synthase.
 DR Pfam; PF00195; Chal_stil_synth; 1.

DR Pfam; PF02797; Chal_stil_synth; 1.
 DR ProDom; PD000453; N-C_synthase; 1.

DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 DR FlavoNoid biosynthesis; Transferase; Acyltransferase;

KW Multigene family.
 FT ACT SITE 164

FT ACT SITE 164 BY SIMILARITY.
 SQ SEQUENCE 389 AA; 42677 MW; 0CC81302CCA8A56A CRC64;

Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAKMPWIOPKTV 18
 |||:|||||:
 DB 111 OKAIKEMGQPKSKI 124

Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAKMPWIOPKTV 18
 |||:|||||:
 DB 111 OKAIKEMGQPKSKI 124

RESULT 14

CHSE_PHANI STANDARD; PRT; 389 AA.
 AC 022046;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase E (EC 2.3.1.74) (Naringenin-chalcone synthase E)
 DE (CHS-E).
 GN CHSE.

OS Pharbitis nil (Violet) (Japanese morning glory).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Convolvulaceae; Ipomoea.

OC NCBI_TaxID=35883;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. KK/ZSK-2; Tissue=Flower buds;
 RX MEDLINE=97393496; PubMed=9249990;

RA Fukuda-Tanaka S., Hoshino A., Hisatomi Y., Habu Y., Hasebe M.,
 RA Iida S.;

RT Identification of new chalcone synthase genes for flower pigmentation
 in the Japanese and common morning glories.";

RL Plant Cell Physiol. 38:754-758(1997).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
 which can under specific conditions spontaneously isomerize into
 naringenin.

CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 OF WHICH ARE BRIGHTLY COLORED.

CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AB001819; BAA21788.1; -
 CC InterPro; IPR001099; N-C synthase.

DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.

DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.

DR FlavoNoid biosynthesis; Transferase; Acyltransferase;

KW Multigene family.
 FT ACT SITE 164

FT ACT SITE 164 BY SIMILARITY.
 SQ SEQUENCE 389 AA; 42685 MW; 0FC0E014B9CC0312 CRC64;

Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAKMPWIOPKTV 18
 |||:|||||:
 DB 111 OKAIKEMGQPKSKI 124

RESULT 15

CHSU_PETHY STANDARD; PRT; 389 AA.
 AC P22928;
 DT 01-APR-1991 (Rel. 19, Created)
 DT 01-APR-1991 (Rel. 25, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase J (EC 2.3.1.74) (Naringenin-chalcone synthase J).
 DE CHSU.

OS Pharbitis nil (Violet) (Japanese morning glory).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; Lamiales; Solanales; Convolvulaceae; Ipomoea.

OC NCBI_TaxID=35883;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. KK/ZSK-2; Tissue=Flower buds;
 RX MEDLINE=97393496; PubMed=9249990;

RA Fukuda-Tanaka S., Hoshino A., Hisatomi Y., Habu Y., Hasebe M.,
 RA Iida S.;

RT Identification of new chalcone synthase genes for flower pigmentation
 in the Japanese and common morning glories.";

RL Plant Cell Physiol. 38:754-758(1997).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
 which can under specific conditions spontaneously isomerize into
 naringenin.

CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 OF WHICH ARE BRIGHTLY COLORED.

CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AB001819; BAA21788.1; -
 CC InterPro; IPR001099; N-C synthase.

DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.

DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.

DR FlavoNoid biosynthesis; Transferase; Acyltransferase;

KW Multigene family.
 FT ACT SITE 164

FT ACT SITE 164 BY SIMILARITY.
 SQ SEQUENCE 389 AA; 42685 MW; 0FC0E014B9CC0312 CRC64;

Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAKMPWIOPKTV 18
 |||:|||||:
 DB 111 OKAIKEMGQPKSKI 124

OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 OC NCBI_TaxID=4102;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Violet 30; TISSUE=leaf;
 RX MEDLINE=90034197; PubMed=2806915;
 RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;
 RT "Cloning and molecular characterization of the chalcone synthase
 multigene family of Petunia hybrida."; Gene 81:245-257(1989).
 RL [2]
 RN SEQUENCE OF 71-389 FROM N.A.
 RC STRAIN=cv. White 137; TISSUE=Anther;
 RA van Tunen A.J.;
 RL Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X14597; CAA32737.1; -;
 DR EMBL; X14599; CAA32739.1; -;
 DR PIR; D72821; SYPJCU;
 DR PIR; S18136; S18136;
 DR InterPro; IPR01099; N-C_synthase.
 DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR ProDom; PD000453; N-C synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 KW Flavonoid biosynthesis; Transferase; Acyltransferase;
 KW Multigene family.
 FT ACT_SITE 164
 FT ACT_SITE 164
 FT CONFLICT 75 E -> V (IN REF. 2).
 SO SEQUENCE 389 AA; 42558 MW; F2B3CDD826FDE7D CRC64;

Query Match 43.7%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 1.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QKAKPWIOPKTV 18
 |||:|||||:
 Db 111 QKAIKENGQPKSKI 124

Search completed: July 30, 2003, 16:24:28
 Job time : 6.875 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 13.375 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-2

Perfect score: 119
Sequence: 1 VYQHGXMKKPMWTPKTKVTPY 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119	100.0	222	1 KABS02	alpha-s2-casein pr
2	95	79.8	223	1 KABS02	alpha-s2-casein pr
3	95	79.8	223	2 UN0547	alpha-s2-casein pr
4	95	79.8	223	2 I46995	alpha-s2-casein C
5	59.5	50.0	235	2 A48383	alpha-s2-casein -
6	59	49.6	182	2 S39776	alpha-s2-casein fo
7	52	43.7	319	2 S18136	naringenin-chalcon
8	52	43.7	340	2 T10957	naringenin-chalcon
9	52	43.7	389	1 SYRUCJ	naringenin-chalcon
10	52	43.7	389	1 SYRUCJ	naringenin-chalcon
11	52	43.7	389	1 SYRUCJ	naringenin-chalcon
12	52	43.7	389	1 SYRUCJ	naringenin-chalcon
13	52	43.7	389	1 SYRUCJ	naringenin-chalcon
14	52	43.7	389	1 SYRUCJ	naringenin-chalcon
15	51	42.9	180	1 S39775	hypothetical prote
16	50	42.0	311	2 T15997	hypothetical prote
17	50	42.0	1877	2 T40550	probable phosphat
18	49	41.2	188	2 JC4680	vascular endotheli
19	49	41.2	207	2 JC4679	hypothetical prote
20	49	41.2	342	2 AD2032	naringenin-chalcon
21	49	41.2	398	2 S16275	naringenin-chalcon
22	49	41.2	400	1 SYZMCC	naringenin-chalcon
23	48	40.3	392	2 S16206	stilbene synthase
24	48	40.3	392	2 S11044	stilbene synthase
25	48	40.3	392	2 S53313	stilbene synthase
26	48	40.3	392	2 S53314	stilbene synthase
27	48	40.3	392	2 S53314	stilbene synthase
28	47	39.5	53	2 S41957	phor-related prote
29	47	39.5	128	2 S33611	naringenin-chalcon

30	47	39.5	158	2 B86169	hypothetical prote
31	47	39.5	285	2 S44369	naringenin-chalcon
32	47	39.5	331	2 J01071	naringenin-chalcon
33	47	39.5	370	2 S44368	naringenin-chalcon
34	47	39.5	383	2 S35165	naringenin-chalcon
35	47	39.5	388	1 SYSCN	naringenin-chalcon
36	47	39.5	388	1 SYSCN	naringenin-chalcon
37	47	39.5	388	1 SYSCN	naringenin-chalcon
38	47	39.5	388	1 SYSCN	naringenin-chalcon
39	47	39.5	388	2 J02249	naringenin-chalcon
40	47	39.5	388	2 J02259	naringenin-chalcon
41	47	39.5	388	2 S60472	naringenin-chalcon
42	47	39.5	388	2 SYSCN	naringenin-chalcon
43	47	39.5	389	1 SYRUCJ	naringenin-chalcon
44	47	39.5	389	1 SYRUCJ	naringenin-chalcon
45	47	39.5	389	2 S20932	naringenin-chalcon

ALIGNMENTS

RESULT 1
KABS02
alpha-s2-casein precursor - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 01-Sep-1981 #sequence revision 03-Feb-1994 #ext_change 22-Jun-1999
C/Accession: J02008; A29087; A91438; S66626; A03107
R/Groenen, M.A.M.; Dijkhof, R.J.M.; Verstege, A.J.M.; van der Poel, J.J.
Gene 123, 187-193, 1993
A>Title: The complete sequence of the gene encoding bovine alpha-s2-casein.
A/Reference number: J02008; MUID:93154583; PMID:8428658
A/Accession: J02008
A>Status: translation not shown
A:Molecule type: DNA
A/Residue: 1-222 <GRO>
A/Cross-references: GB:M94327
R/Stewart, A.F.; Bomsing, J.; Beattie, C.W.; Shah, F.; Willis, I.M.; Mackinlay, A.G.
Mol. Biol. Evol. 4, 231-241, 1987
A>Title: Complete nucleotide sequences of bovine alpha-s2- and beta-casein cDNAs: compar
A/Reference number: A93062; MUID:18818989; PMID:2833669
A/Accession: A29087
A>Status: translation not shown
A:Molecule type: mRNA
A/Residue: 1-222 <STB>
A/Cross-references: GB:M16644; NID:9162928; PTDN:AA30479.1; PTD:9162929
R/Brignon, G.; Ribadeau Dumas, B.; Mercier, J.C.; Pelissier, J.P.; Dae, B.C.
FEBS Lett. 76, 274-279, 1977
A>Title: Complete amino acid sequence of bovine alpha-s2-casein.
A/Reference number: A91438; MUID:77185633; PMID:862906
A/Accession: A91438
A:Molecule type: protein
A/Residue: 16-101, 'EE', 104-222 <BRI>
A/Note: Four fractions, previously designated s2, s3, s4, and s6, appear to have the sam
these
J. Dairy Res. 46, 211-213, 1979
A>Title: A genetic and biochemical analysis of a polymorphism of bovine alpha-s2-casein
A/Reference number: A92771; MUID:79239637; PMID:469044
A/Contents: annotation; D allele
A/Note: the sequence of the D allele has a deletion of nine residues, which may be 49-58
R/Zucht, H.D.; Raide, M.; Adrmann, K.; Maegert, H.J.; Forsmann, W.G.
FEBS Lett. 372, 185-188, 1995
A>Title: Casein alpha-s2 derived peptide exhibits antibacterial activity
A/Reference number: S66626; MUID:9600204; PMID:755666
A/Accession: S66626
A:Molecule type: protein
A/Residue: 165-203 <ZUC>
C/Comment: The sequence of the A allele is shown.
C/genetics:
A/Gene: alphaS2ca
A/Map position: 6
A/Introns: 17/3; 26/3; 33/3; 47/3; 56/3; 65/3; 74/3; 82/3; 97/3; 138/3; 147/3; 156/3; 16

C:Superfamily: alpha-s2-casein
 C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-223/Product: alpha-s2-casein #status experimental <MAT>
 F:23,24,25,31,71,72,73,76,144,146,158/Binding site: phosphate (Ser) (covalent) #status P

Query Match 100.0%; Score 119; DB 1; Length 223;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWIOPKTKVIPPY 21
 DB 198 VYOHQKMKPWIOPKTKVIPPY 218

RESULT 2

alpha-s2-casein precursor - sheep
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C>Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 22-Jun-1999
 C:Accession: A25070; S17856
 R:Boisnard, M.; Pectisnant, G.
 Biochimie 67, 1043-1051, 1985
 A>Title: Complete sequence of ovine alpha-s2-casein messenger RNA.
 A:Reference number: A25070; MUID:86104467; PMID:3002499
 A:Accession: A25070
 A:Molecule type: mRNA
 A:Residues: 1-223 <BO1>

A:Cross-references: GB:X03238; NID:g1238; PIDN:CAA6983.1; PID:g732894
 A:Note: 64 Aam was also found
 R:Boisnard, M.; Hue, D.; Boumlo, C.; Mercier, J.C.; Gaye, P.
 Eur. J. Biochem. 201, 633-641, 1991
 A>Title: Multiple mRNA species code for two non-allelic forms of ovine alpha-s2-casein.
 A:Reference number: S17856; MUID:92036419; PMID:1935959

A:Accession: S17856
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-93, 'T', '95-223 <BO2>
 C:Superfamily: alpha-s2-casein
 C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-223/Product: alpha-s2-casein #status predicted <KA2>
 F:23,24,25,32,55,72,73,74,77,145,147,159/Binding site: phosphate (Thr) (covalent) #status
 F:53,86,146,154,170,198/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 79.8%; Score 95; DB 1; Length 223;
 Best Local Similarity 81.0%; Pred. No. 5.5e-07;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWIOPKTKVIPPY 21
 DB 199 VYOHQKMKPWIOPKTKVIPPY 219

RESULT 3

alpha-s2-casein precursor - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C>Date: 31-Dec-1993 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: S33881; S33880; JN0547; S20620
 R:Boumlo, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Protein Seq. Data Anal. 5, 213-218, 1993
 A>Title: Characterization of goat allelic alpha-s2-caseins A and B: further evidence of t

A:Accession: S33881
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-223 <BOU>
 A:Note: allele B
 A:Accession: S33880
 A:Molecule type: protein
 A:Residues: 16-78, 'E', 80-223 <BO2>
 A:Note: sequence deduced from compositional analysis of peptides

R:Boumlo, C.
 Gene 125, 235-236, 1993
 A>Title: Sequence of the goat alpha-s2-casein-encoding cDNA.
 A:Reference number: JN0547; MUID:93216130; PMID:8462880

A:Accession: JN0547
 A:Molecule type: mRNA
 A:Residues: 1-78, 'E', 80-223 <BO3>
 A:Cross-references: EMBL:X65160; NID:g955; PIDN:CAA6278.1; PID:g9556
 A:Note: allele A

C:Superfamily: alpha-s2-casein
 C:Keywords: mammary gland; milk; phosphoprotein
 F:1-15/Domain: signal sequence #status predicted <SIG>
 F:16-223/Product: alpha-s2-casein #status experimental <MAT>
 F:23,24,25,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status
 F:53,146/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 79.8%; Score 95; DB 2; Length 223;
 Best Local Similarity 81.0%; Pred. No. 5.5e-07;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWIOPKTKVIPPY 21
 DB 199 VYOHQKMKPWIOPKTKVIPPY 219

RESULT 4

alpha-s2-casein C - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C>Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 13-Aug-1999
 C:Accession: I46995
 R:Boumlo, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Anim. Genet. 25, 173-177, 1994

A>Title: Biochemical and genetic analysis of variant C of caprine alpha-s2-casein (Capr
 A:Reference number: I46995; MUID:95030556; PMID:7943951
 A:Accession: I46995
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-223 <BOU>
 A:Cross-references: GB:S74171; NID:g707033; PIDN:AAB32166.1; PID:g707034
 C:Superfamily: alpha-s2-casein

Query Match 79.8%; Score 95; DB 2; Length 223;
 Best Local Similarity 81.0%; Pred. No. 5.5e-07;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWIOPKTKVIPPY 21
 DB 199 VYOHQKMKPWIOPKTKVIPPY 219

RESULT 5

alpha-s2-casein - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
 C:Accession: A48383
 R:Alexander, L.J.; Das Gupta, N.A.; Beattie, C.W.
 Anim. Genet. 23, 365-367, 1992

A>Title: The sequence of porcine alpha-s2-casein cDNA.
 A:Reference number: A48383; MUID:92367960; PMID:1503576
 A:Accession: A48383
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-235 <ALB>
 A:Experimental source: mammary gland
 A:Note: sequence inconsistent with the nucleotide translation
 A:Note: sequence extracted from NCBI backbone (NCBIN:110884, NCBIP:110885)
 C:Superfamily: alpha-s2-casein

Query Match 50.0%; Score 59.5; DB 2; Length 235;
 Best Local Similarity 56.5%; Pred. No. 0.15;
 Matches 13; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

```

QY      1 VYQHQAAMKRPWIQPKT--KVIP 20
      |||:|||||  |||
DB      208 VHQAQAAMKRPWNHIKTNSYQIIP 230

```

RESULT 6
S39776
alpha-S2-casein form b precursor - rabbit
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 13-Aug-1999
R/Accession: S39776
R/Dawson, S.P.; Wilde, C.J.; Tighe, P.J.; Mayer, R.J.
Biochem. J. 296, 777-784, 1993
A/Title: Characterization of two novel casein transcripts in rabbit mammary gland
A/Reference number: S39775; MUID:94107245; PMID:8280077
A/Accession: S39776
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-182 <DMW>
A/Cross-references: EMBL:X76909; NID:g439527; PIDD:CAA54231.1; PID:g439528
C/Superfamily: alpha-S2-casein

Query Match	49.6%;	Score 59;	DB 2;	Length 182;
Best Local Similarity	42.9%;	Pred. No. 0.13;		
Matches	9;	Conservative	5;	Mismatches 7;
				Indels 0;
				Gaps 0

```
QY      1 VYQHQAAMKRPWIGPRTKVIPIY 21
      :||: ||| : ||| ||:
Db      63 LYQYPTVMDPWTRAETKAIRP 103
```

```

RESULT 7
S18136
naringenin-chalcone synthase (EC 2.3.1.74) - garden petunia
C/Species: Petunia x hybrida (garden petunia)
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-May-2000
C/Accession: S18136
R/van Tunen, A.J.
submitted to the EMBL Data Library, March 1989
A/Reference number: S18136
A/Accession: S18136
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-319 <TRN>
A/Cross-references: EMBL:X14599; NID:g20537; PID:CA32739.1; PID:g20538
C/Superfamily: chalcone synthase
C/Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

```

Query Match	43.7%	Score 52;	DB 2;	Length 319;
Best Local Similarity	64.3%	Pred. No. 2.8;		
Matches	9;	Conservative	3;	Mismatches 2;
				Indels 0;
				Gaps 0

QY	5	QKAMKPWIQPKTKV	18
		: : :	
Db	41	QKAIKEWGQPKSKI	54

RESULT 8
T10957
naringenin-chalcone synthase (EC 2.3.1.74) CHS-FL1 - common morning-glory
C/Species: Ipomoea purpurea (common morning-glory)
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 05-May-2000
C/Accession: T10957
R/Rausher, M.D.; Tiffin, P.L.; Miller, R.R.
submitted to the EMBL Data Library, January 1997
A/Description: Regulation of anthocyanin gene expression in Ipomoea purpurea
A/Reference number: Z17229
A/Accession: T10957
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-340 <RAU>
A/Cross-references: EMBL:U74082; NID:G1658119; PTD:G1658120

A: Experimental source: flower bud; WMAA genotype (dark flowered)
C: Genetics:
A: Gene: CHS-FL1
C: Superfamily: chalcone synthase
C: Keywords: acyltransferase; coenzyme A

Query Match	43.7%	Score 52;	DB 2;	Length 340;
Best Local Similarity	64.3%	Pred. No. 3;		
Matches	9;	Conservative	3;	Mismatches 2;
				Indels 0;
				Gaps 0

QY	5	QKAMKPWIQPKTKV	18
		: :	
Db	111	QKAIKEWGQPKSKI	124

RESULT 9
 SYRJCJ
 naringenin-chalcone synthase (EC 2.3.1.74) J - garden petunia
 N:Alternate names: chalcone synthase
 C:Species: Petunia x hybrida (garden petunia)
 C:Date: 30-Sep-1991 #sequence_revision 17-Mar-2000 #text_change 05-May-2000
 C:Accession: D72821; J80309
 R:Koes, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.
 Gene 81, 245-257, 1989
 A:Title: Cloning and molecular characterization of the chalcone synthase multi-gene family
 A:Reference number: J80308; MUID:J90034197; PMID:2806915
 A:Accession: D72821
 A:Molecule type: DNA
 A:Residues: 1-389 <KOS>
 A:Cross-references: EMBL:X14597; NID:g20535; PIDN:CMA32737.1; PID:g20536
 A:Experimental source: strain Violet 30, leaf

Query Match	43.7%	Score 52;	DB 1;	Length 389;
Best Local Similarity	64.3%	Pred. NO. 3.4;		
Matches	9;	Conservative	3;	Mismatches
			2;	Indels
				Gaps
				0

QY 5 QKAMKPWIOPTKV 18
|||:| |||:|:
Db 111 QKAIKEWGP KSKI 124

```

RESULT 10
SYNCH
naringenin-chalcone synthase (EC 2.3.1.74) R - garden petunia
N/Alternate names: chalcone synthase
C/Species: Petunia x hybrida (garden petunia)
C/Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 05-May-2000
C/Accession: A23643
R/Koss, R.E.; Spelt, C.E.; Reif, H.U.; van den Elzen, P.J.M.; Veltkamp, E.; Mol, J.N.M.
Nucleic Acids Res. 14, 5229-5239, 1986
A/Title: Floral tissue of Petunia hybrida (V30) expresses only one member of the chalcon
A/Reference number: A23643; MUID:86286540; PMID:3016642
A/Accession: A23643
A/Molecule type: mRNA
A/Residues: 1-389 <KO>
A/Cross-references: GB:X04080; NID:G20541; PIDN:CAA2718.1; PID:G20542
A/Experimental source: strain Violet 30, flowers
C/Comment: This enzyme plays a central role in the biosynthesis of all classes of flavon
C/Genetics:
A:Gene: chcr
A:Note: expressed in floral tissue
C:Superfamily: chalcone synthase

```

C;Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 43.7%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKEMGQPKSKI 124

RESULT 11

STPCCA

naringenin-chalcone synthase (EC 2.3.1.74) A - garden pecunia

N;Alternate names: chalcone synthase

C;Species: *Petunia x hybrida* (garden petunia)

C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000

C;Accession: J50308

R;Koei, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.

Gene 81, 245-257, 1989

A;Title: Cloning and molecular characterization of the chalcone synthase multigene family

A;Reference number: J50308; PMID:90034197; PMID:2806915

A;Accession: J50308

A;Molecule type: DNA

A;Residues: 1-389 <KOE>

A;Cross-references: GB:X14591; NID:G20524; PIDN:CAA32731.1; PID:G20525

A;Experimental source: strain Violet 30, leaf

C;Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoid

C;Genetics:

A;Gene: chsA

A;Map position: V

A;Note: chsA is the major expressed member of the gene family in various floral tissues

C;Superfamily: chalcone synthase

C;Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 43.7%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKEMGQPKSKI 124

RESULT 12

JC5136

naringenin-chalcone synthase (EC 2.3.1.74) 2 - potato

N;Alternate names: chalcone synthase; CHS

C;Species: *Solanum tuberosum* (potato)

C;Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 05-May-2000

C;Accession: J51316; PC4239

R;Jedon, J.H.; Kim, H.S.; Choi, K.H.; Joung, Y.H.; Joung, H.; Byun, S.M.

Biosect. Biotechnol. Biochem. 60, 1907-1910, 1996

A;Title: Cloning and characterization of one member of the chalcone synthase gene family

A;Reference number: J51316; PMID:97141614; PMID:8987872

A;Accession: J51316

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-389 <JEO1>

A;Cross-references: GB:U47738; NID:G1470059; PIDN:AAB05239.1; PID:G1470060

A;Accession: PC4239

A;Status: preliminary

A;Molecule type: protein

A;Residues: 158-165;167-373 <JEO2>

C;Comment: This enzyme is important in the biosynthesis of all classes of flavonoids in

C;Superfamily: chalcone synthase

C;Keywords: acyltransferase; coenzyme A

Query Match 43.7%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTV 18

DB 111 OKAIKEMGQPKSKI 124

RESULT 13

T07799

naringenin-chalcone synthase (EC 2.3.1.74) - common morning-glory

N;Alternate names: chalcone synthase

C;Species: *Ipomoea purpurea* (common morning-glory)

C;Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000

C;Accession: T07799

R;Fukada-Tanaka, S.; Hoshino, A.; Hisatomi, Y.; Habu, Y.; Hasebe, M.; Iida, S.

Plant Cell Physiol. 38, 754-758, 1997

A;Title: Identification of new chalcone synthase genes for flower pigmentation in the *Ipomoea*

A;Reference number: Z16140; PMID:97393496; PMID:9249990

C;Accession: T07799

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-389 <PDK>

A;Cross-references: EMBL:AB001827; NID:G2329836; PIDN:BA21789.1; PID:G2329837

C;Genetics:

A;Gene: CHD-B

C;Superfamily: chalcone synthase

C;Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 43.7%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKEMGQPKSKI 124

RESULT 14

SYKCD

naringenin-chalcone synthase (EC 2.3.1.74) - garden snapdragon

N;Alternate names: chalcone synthase

C;Species: *Antirrhinum majus* (garden snapdragon)

C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000

C;Accession: S07312; A33217

R;Sommer, H.; Saedler, H.

Mol. Gen. Genet. 202, 429-434, 1986

A;Title: Structure of the chalcone synthase gene of *Antirrhinum majus*.

A;Reference number: S07312

A;Accession: S07312

A;Molecule type: DNA

A;Residues: 1-390 <SOM>

A;Cross-references: EMBL:X03710; NID:G16015; PIDN:CAA27338.1; PID:G16016

A;Accession: A33217

A;Status: translation not shown

A;Molecule type: mRNA

A;Residues: 1-390 <SOM2>

C;Genetics:

A;Gene: chs

A;Introns: 60/1; 162/3

C;Superfamily: chalcone synthase

C;Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 43.7%; Score 52; DB 1; Length 390;
Best Local Similarity 64.3%; Pred. No. 3.4;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKEMGQPKSKI 124

RESULT 15

S39775

alpha-92-casein form a precursor - rabbit

C;Species: *Oryctolagus cuniculus* (domestic rabbit)

C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 13-Aug-1999

C:Accession: S39775
 R: Dawson, S. P.; Wilde, C. J.; Tighe, P. J.; Mayer, R. J.
 Biochem. J. 296, 777-784, 1993
 A: Title: Characterization of two novel casein transcripts in rabbit mammary gland
 A: Reference number: S39775; MUID: 94107245; PMID: 8280077
 A: Accession: S39775
 A: Status: preliminary
 A: Molecule type: mRNA
 A: Residues: 1-180 <DAW>
 A: Cross-references: EMBL: X76907; NID: g439525; PIDN: CAA54228.1; PID: g439526
 C: Superfamily: gamma-casein

Query Match	42.9%;	Score 51;	DB 2;	Length 180;
Best Local Similarity	50.0%;	Pred. No. 2.2;		
Matches 10; Conservative	0;	Mismatches 10;	Indels 0;	Gaps 0;

```

QY      1 VYQHQAAMKRWIQPKTKVIP 20
          |||  |||  |||  |||
DB      99 VYQQQIVMNPWKHKVKTTP 118

```

Search completed: July 30, 2003, 16:31:15
Job time : 14.375 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 41.375 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-2

Perfect score: 119
Sequence: 1 VYOHQKAMKPMIOPKTVIPY 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_15Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	100.0	21	21	AA198338
2	119	100.0	24	15	AA1980481
3	119	100.0	24	21	AA198339
4	119	100.0	25	13	AA1980055
5	119	100.0	25	21	AA198340
6	119	100.0	31	18	AA1982219
7	119	100.0	31	23	AA197466
8	119	100.0	222	18	AA1982220
9	119	100.0	222	23	AA197468

10	108.5	91.2	26	20	AA193885	Blifidobacterium bi
11	95	79.8	223	23	AA197469	Alpha-S2 casein pr
12	95	79.8	223	23	AA197470	Goat alpha-S2 case
13	95	79.8	223	23	AA197471	Capra hircus alpha
14	95	79.8	223	23	AA197473	Sheep alpha-S2 cas
15	95	79.8	223	23	AA197475	Alpha-S2 casein pr
16	92	77.3	16	21	AA198337	Internal peptide f
17	85	71.4	19	18	AA198218	Alpha-S2 casein pr
18	85	71.4	19	23	AA197465	Bovine alpha-S2 ca
19	59.5	50.0	214	23	AA197474	Pig alpha-S2 casei
20	59.5	49.6	178	23	AA197472	Rabbit alpha-S2 ca
21	53	44.5	9	17	AA1980482	Milk derived anti-
22	52	43.7	360	22	AA1960169	Petunia chalcone s
23	52	43.7	389	17	AA196125	Chalcone synthase.
24	52	43.7	389	22	AA1974019	Cyclamen chalcone
25	52	43.7	389	24	AA1973901	Petunia chalcone s
26	50	42.0	291	22	AA1969992	Mouse GN protein s
27	49	41.2	8	17	AA1980478	Milk derived anti-
28	49	41.2	13	23	AA197464	Bovine alpha-S2 ca
29	49	41.2	133	17	AA1980482	Vascular endotheli
30	49	41.2	133	20	AA1980492	Murine vascular en
31	49	41.2	133	23	AA1983406	Murine endotheli
32	49	41.2	160	20	AA1986208	Human VEGF-B trunc
33	49	41.2	167	20	AA1986234	Human VEGF-B full
34	49	41.2	188	17	AA1980486	Heart vascular en
35	49	41.2	188	17	AA1980486	Murine VEGF167. Mu
36	49	41.2	188	20	AA1986201	Human vascular en
37	49	41.2	188	20	AA1980490	Murine vascular en
38	49	41.2	188	23	AA1983404	Murine vascular en
39	49	41.2	195	17	AA1980482	Heart vascular en
40	49	41.2	195	20	AA1980491	Murine vascular en
41	49	41.2	195	23	AA1983405	Vascular endotheli
42	49	41.2	207	17	AA1980483	Vascular endotheli
43	49	41.2	207	17	AA1980483	Murine VEGF166. Mu
44	49	41.2	207	20	AA1980494	Murine vascular en
45	49	41.2	207	22	AA19836296	Human vascular en

ALIGNMENTS

RESULT 1
AA198338
ID AA198338 standard; peptide; 21 AA.
AC AA198338:
XX
DT 14-JUL-2000 (first entry)
XX
DE Internal peptide fragment of bovine alpha-S2 casein #2.
XX
KW Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
XX
OS Bos sp.
XX
PN WO200015655-A1.
XX
PD 23-MAR-2000.
XX
PF 15-SEP-1999; 99WO-EP07002.
XX
PR 15-SEP-1998; 98EP-0203107.
PR 08-JUN-1999; 99EP-0201815.
XX
PA (NIZO-) NIZO FOOD RES.
XX
PI Visser S, Recto I;
XX
DR WPI, 2000-271377/23.
XX
PT Novel process for producing peptides with e.g. antimicrobial activity
from biological fluids such as milk, whey or blood comprises contacting

PT fluid with chromatographic medium to adsorb peptide domain of interest
 PT
 XX
 PS
 XX
 XX
 Claim 14; Page 22; 41pp; English.

CC This sequence represents an internal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

SO Sequence 21 AA;

Query Match 100.0%; Score 119; DB 21; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKAMKPMIOPKTKVIPPY 21
 DB 1 VYOHQKAMKPMIOPKTKVIPPY 21

RESULT 2

ID AAR60481 standard; peptide; 24 AA.

AC AAR60481;

DT 19-APR-1995 (first entry)

XX Peptide used in tranquilliser.

KW Peptide; tranquilliser; food; foodstuff; mental stabilisation.

OS Synthetic.

PN JP06211689-A.

PD 02-AUG-1994.

PF 19-JAN-1993; 93JP-0024811.

PR 19-JAN-1993; 93JP-0024811.

XX (KANE) KANEBO LTD.

DR WPI; 1994-283276/35.

PT Synthetic peptide(s) used in tranquilliser - also used in the
 production of a food for mental stabilisation.

PS Claim 3; Page 1; 5pp; Japanese.

CC The peptide may be used as a component of a tranquilliser which may
 CC be administered orally and is low in toxicity. The peptide may also
 CC be used as a component of a foodstuff which may be used for mental
 CC stabilisation. See also AAR60480.

SO Sequence 24 AA;

Query Match 100.0%; Score 119; DB 15; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.7e-10;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKAMKPMIOPKTKVIPPY 21
 DB 1 VYOHQKAMKPMIOPKTKVIPPY 21

RESULT 3

ID AAY88339 standard; peptide; 24 AA.

AC AAY88339;

DT 14-JUL-2000 (first entry)

DE Internal peptide fragment of bovine alpha-S2 casein #3.

KW Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
 KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.

OS Bos sp.

PN WO200015655-A1.

PD 23-MAR-2000.

PF 15-SEP-1999; 99WO-EP07002.

PR 15-SEP-1998; 98EP-0203107.

PR 08-JUN-1993; 99EP-0201815.

PA (NIZO-) NIZO FOOD RES.

PI Visser S, Recto I;

DR WPI; 2000-271377/23.

PT Novel process for producing peptides with e.g. antimicrobial activity
 from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest

XX Claim 14; Page 22; 41pp; English.

CC This sequence represents an internal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

SO Sequence 24 AA;

Query Match 100.0%; Score 119; DB 21; Length 24;
 Best Local Similarity 100.0%; Pred. No. 1.7e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKAMKPMIOPKTKVIPPY 21
 DB 1 VYOHQKAMKPMIOPKTKVIPPY 21

RESULT 4

AA20055
ID AAR20055 standard; peptide; 25 AA.
XX
AC AAR20055;
XX
DT 25-MAR-2003 (updated)
DT 26-MAR-1992 (first entry)
XX
DE Casein peptide.
XX
KW Casein; platelet; aggregation; thrombosis; collagen.
XX
OS Synthetic.
XX
PN JP03255095-A.
XX
PD 13-NOV-1991.
XX
PF 02-MAR-1990; 90CP-0052554.
XX
PR 02-MAR-1990; 90CP-0052554.
XX
PA (KANE) KANEBO LTD.
XX
DR WPI; 1992-002669/01.
XX
PT Casein peptide(s) for treating thrombosis - as inhibitors of
PT platelet aggregation caused by adenosine-5'-di-phosphate and
PT collagen and as biochemical reagents
XX
PS Claim 1; Page 1; 8pp; Japanese.
XX
CC Leu25 may be omitted. The peptide and its salts have inhibiting
CC activity against platelet aggregation caused by adenosine-5'-
CC di-phosphate and by collagen. They are useful for prevention and
CC treatment of thrombosis and are also useful as a biochemical reagent.
CC (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 25 AA;
XX
Query Match 100.0%; Score 119; DB 13; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VYOHOKAMKPMIOPKTKVIPPY 21
DB 1 VYOHOKAMKPMIOPKTKVIPPY 21
XX
RESULT 5
AA20055
ID AAY88340 standard; peptide; 25 AA.
XX
AC AAY88340;
XX
DT 14-JUL-2000 (first entry)
XX
DE C-terminal peptide fragment of bovine alpha-S2 caesin.
XX
KW Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;
KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
XX
OS Bos sp.
XX
PN WO200015655-A1.
XX
PD 23-MAR-2000.
XX
PF 15-SEP-1999; 99WO-EP07002.
XX
PR 15-SEP-1998; 98EP-0203107.
PR 08-JUN-1999; 98EP-0201815.
XX
PT

PA (NIZO-) NIZO FOOD RES.
XX
PI Visser S, Recio I;
XX
DR WPI; 2000-271377/23.
XX
XX Novel process for producing peptides with e.g. antimicrobial activity
PT from biological fluids such as milk, whey or blood comprises contacting
PT fluid with chromatographic medium to adsorb peptide domain of interest
PT
XX
PS Claim 14; Page 22; 41pp; English.
XX
XX This sequence represents a C-terminal fragment of bovine alpha-S2 caesin
CC protein. The peptide is an example of a peptide with antibacterial
CC activity that can be produced by the process of the invention. The
CC invention relates to a process for producing peptides from biological
CC fluids. The process comprises chromatography of the biological fluid, in
CC situ hydrolysis of selectively bound peptides, washing to remove unbound
CC peptide, and elution of the peptides of interest. The process is used for
CC producing peptides from biological fluids, such as milk, whey or blood.
CC For example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumour activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptides with a
CC selected activity of interest without the need for intermediate
CC purification of the precursor protein.
XX
SQ Sequence 25 AA;
XX
Query Match 100.0%; Score 119; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VYOHOKAMKPMIOPKTKVIPPY 21
DB 1 VYOHOKAMKPMIOPKTKVIPPY 21
XX
RESULT 6
AA20055
ID AAM32219 standard; peptide; 31 AA.
XX
AC AAM32219;
XX
DT 03-FEB-1998 (first entry)
XX
DE Alpha-S2 caesin precursor C-terminal peptide fragment #5.
XX
KW Alpha-S2 caesin precursor; growth promoting; mitogenic assay;
KW platelet-derived growth factor; insulin-like derived growth factor;
XX
OS Synthetic.
XX
OS Bos taurus.
XX
PN WO9716460-A1.
XX
PD 09-MAY-1997.
XX
PF 31-OCT-1996; 96WO-GB02658.
XX
PR 31-OCT-1995; 95GB-0022302.
XX
PA (UYLI-) UNIV LIVERPOOL.
XX
PI Liu Q, Smith JA, Wilkinson MC;
XX
DR WPI; 1997-272048/24.
XX
PT Manufacture of medicament or foodstuff for promoting growth - using
PT peptide(s) with a sequence identical to the C-terminal end of an

PT alpha-S2 casein precursor
 XX
 PS Claim 12; Page 21; 33pp; English.
 XX
 CC The present sequence, which is substantially identical to the C-terminal
 CC end of an alpha-S2 casein precursor, was found after storage in PBS to
 CC exhibit growth promoting activity for rat mammary fibroblast cell line
 CC (Rama 27), which is not significantly stimulated by platelet-derived
 CC growth factor or insulin-like growth factor. The activity of the peptide
 CC increased when maintained at alkaline pH. By way of contrast, alpha-S2
 CC casein was inactive in a mitogenic assay. This peptide may be used in the
 CC manufacture of a medicament or foodstuff for promoting growth in
 CC humans or animals.
 CC
 SQ Sequence 31 AA;
 XX
 Query Match 100.0%; Score 119; DB 18; Length 31;
 Best Local Similarity 100.0%; Pred. No. 2.2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIOPKTKVIRY 21
 DB 7 VYQHOKAMKPMIOPKTKVIRY 27
 XX
 RESULT 7
 AAE17466
 ID AAE17466 standard; peptide; 31 AA.
 XX
 AC AAE17466;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor protein C-terminal fragment #7.
 XX
 KW Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KW chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Bos sp.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 Claim 9; Page 22; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptide stimulates the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is a peptide
 CC derived from the C-terminal of bovine alpha-S2 casein precursor protein.

XX
 SQ Sequence 31 AA;
 XX
 Query Match 100.0%; Score 119; DB 23; Length 31;
 Best Local Similarity 100.0%; Pred. No. 2.2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIOPKTKVIRY 21
 DB 7 VYQHOKAMKPMIOPKTKVIRY 27
 XX
 RESULT 8
 AAW32220
 ID AAW32220 standard; protein; 222 AA.
 XX
 AC AAW32220;
 XX
 DT 03-FEB-1998 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor.
 XX
 KW Bovine alpha-S2 casein precursor; growth promoting; mitogenic assay;
 KW platelet-derived growth factor; insulin-like derived growth factor;
 XX
 OS Bos taurus.
 XX
 PN WO9716460-A1.
 XX
 PD 09-MAY-1997.
 XX
 PF 31-OCT-1996; 96WO-GB02658.
 XX
 PR 31-OCT-1995; 95GB-0022302.
 XX
 PA (UYLI-) UNIV LIVERPOOL.
 XX
 PI Liu Q, Smith JA, Wilkinson MC;
 XX
 DR WPI; 1997-272048/24.
 XX
 PT Manufacture of medicament or foodstuff for promoting growth - using
 PT peptide(s) with a sequence identical to the C-terminal end of an
 PT alpha-S2 casein precursor
 XX
 PS Disclosure; Page 3; 33pp; English.
 XX
 CC The present sequence represents bovine alpha-S2 casein precursor.
 CC Peptides having an amino acid sequence which is substantially identical
 CC to the C-terminal end of an alpha-S2 casein precursor, are used for
 CC manufacture of a medicament or foodstuff for promoting growth in humans
 CC or animals.
 CC
 SQ Sequence 222 AA;
 XX
 Query Match 100.0%; Score 119; DB 18; Length 222;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMIOPKTKVIRY 21
 DB 198 VYQHOKAMKPMIOPKTKVIRY 218
 XX
 RESULT 9
 AAE17468
 ID AAE17468 standard; protein; 222 AA.
 XX
 AC AAE17468;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor protein.

XX Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Bos sp.
 XX WO200202133-A2.
 PN 10-JAN-2002.
 XX
 PD 13-JUN-2001; 2001WO-GB02601.
 XX
 PF 30-JUN-2000; 2000GB-0016189.
 XX
 PR (PEPS-) PEPSYN LTD.
 XX
 PA Smith JA;
 XX
 PI WPI, 2002-154690/20.
 DR
 XX
 XX Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 XX Claim 8; Page 6; 27pp; English.
 PS
 XX The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is bovine
 CC alpha-S2 casein precursor protein.
 CC
 XX Sequence 222 AA;
 SQ
 Query Match 100.0%; Score 119; DB 23; Length 222;
 Best Local Similarity 100.0%; Pred. No. 1.8e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMWIOPTKVIPIY 21
 DB 198 VYQHOKAMKPMWIOPTKVIPIY 218
 RESULT 10
 AAM93885
 ID AAM93885 standard; peptide; 26 AA.
 AC AAM93885;
 XX
 XX 25-JUN-1999 (first entry)
 DT
 XX Bifidobacterium bifidus stimulating peptide 21.
 DE
 XX Bifidogenic peptide; protease; treatment; microbe-related disease;
 KW bacteria; fungi; Yeast; protozoa; virus; mycoplasma; filaria; plasmodia;
 KM infection; inflammation; microbial induced tumour; degenerative disorder;
 KM diarrhoea; colic; oral microflora; intestinal microflora; caries;
 KM vaginal microflora.
 XX
 XX Bifidobacterium bifidus.
 OS
 XX WO914231-A2.
 PN
 XX 25-MAR-1999.
 PD
 XX

PF 16-SEP-1998; 98WO-EP05899.
 XX
 XX 11-FEB-1998; 98DE-1005385.
 PR 16-SEP-1997; 97DE-1040604.
 XX
 XX (FORS/) FORSMANN W.
 PA
 XX Forsemann W, Liepke C, Zucht H;
 PI WPI, 1999-244022/20.
 DR
 XX
 XX Milk-derived peptides that stimulate Bifidobacterium bifidus
 PT
 PS Claim 2; Page 3; 25pp; German.
 XX
 XX This invention describes milk-derived bifidogenic peptides and their
 CC active derivatives or fragments, and combinations of them produced by
 CC chemical coupling. Such are produced from bovine or human milk by
 CC treatment for 2 hr with proteases, then centrifuging to remove fat and
 CC acidifying to pH 2 to precipitate proteins. The solution phase is then
 CC subjected to reverse-phase high-performance liquid chromatography (HPLC)
 CC and cation-exchange HPLC, the fractions adjusted to salt content below
 CC 25 mM (by dialysis or reverse-phase HPLC) and tested for activity by
 CC growing Bifidobacterium bifidus and Escherichia coli in presence of the
 CC fractions. Those fractions for which (Bw-B0)-(Ew-E0) is at least 0.15
 CC are selected where Bw = germ count after 16 hr culture of B. bifidus in
 CC 50% Bliker broth containing peptide at 0.2 mg/ml, B0 = germ count under
 CC similar conditions in a peptide-free control. Ew = germ count after 16
 CC hr culture of E. coli in 3 g/l tryptic broth containing peptide at
 CC 0.2 mg/ml, E0 = germ count under similar conditions in a peptide-free
 CC control. The peptides AAM93865-W93888 are used to treat microbe-related
 CC diseases caused by bacteria, fungi, Yeast, protozoa, viruses,
 CC mycoplasma, filaria and plasmodia, e.g. infections, inflammation,
 CC microbially induced tumours or degenerative disorders, diarrhoea, colic,
 CC abnormalities in oral, intestinal or vaginal microflora, or caries.
 CC
 XX Sequence 26 AA;
 SQ
 Query Match 91.2%; Score 108.5; DB 20; Length 26;
 Best Local Similarity 95.5%; Pred. No. 6.2e-09;
 Matches 21; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 VYQHOKAMKPMWIOPTKVIPIY 21
 DB 1 VYQHOKAMKPMWIOPTKVIPIY 22
 RESULT 11
 AAE17469
 ID AAE17469 standard; Protein; 223 AA.
 AC AAE17469;
 XX
 XX 22-APR-2002 (first entry)
 DT
 XX Alpha-S2 casein precursor (alpha-S2-CN) protein #1.
 DE
 XX Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
 KW fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
 KM toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 XX Unidentified.
 OS
 XX WO200202133-A2.
 PN
 XX 10-JAN-2002.
 PD
 XX
 XX 13-JUN-2001; 2001WO-GB02601.
 PF
 XX 30-JUN-2000; 2000GB-0016189.
 PR
 XX (PEPS-) PEPSYN LTD.
 PA
 XX

PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Disclosure; Page 8; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is alpha-S2
 CC casein precursor (alpha-S2-CN) protein.
 XX
 SQ Sequence 223 AA;
 XX
 Query Match 79.8%; Score 95; DB 23; Length 223;
 Best Local Similarity 81.0%; Pred. No. 5,7e-06;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 VYOHQKMKPWTQPKTNALPY 21
 DB 199 VDOHQKMKPWTQPKTNALPY 219
 XX
 RESULT 12
 ID AAE17470 standard; Protein; 223 AA.
 XX
 AC AAE17470;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Goat alpha-S2 casein E precursor protein.
 XX
 KM Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC derivative. The peptide contains an amino acid sequence from alpha-S2

CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein E precursor protein.
 XX
 SQ Sequence 223 AA;
 XX
 Query Match 79.8%; Score 95; DB 23; Length 223;
 Best Local Similarity 81.0%; Pred. No. 5,7e-06;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 VYOHQKMKPWTQPKTNALPY 21
 DB 199 VDOHQKMKPWTQPKTNALPY 219
 XX
 RESULT 13
 ID AAE17471 standard; Protein; 223 AA.
 XX
 AC AAE17471;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Capra hircus alpha-S2 casein C precursor protein.
 XX
 KM Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein C precursor protein.
 XX
 SQ Sequence 223 AA;
 XX
 Query Match 79.8%; Score 95; DB 23; Length 223;

Best Local Similarity 81.0%; Pred. No. 5.7e-06;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKAMKRWIOPTKTVIPY 21
DB 199 VDOHQKAMKRWIOPTKTVIPY 219

RESULT 14
AAE17473
ID AAE17473 standard; Protein; 223 AA.

AC AAE17473;

DT 22-APR-2002 (first entry)

DE Sheep alpha-S2 casein precursor protein.

KW Sheep; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
chewing gum; cosmetic; wrinkling; periodontal disease.

OS Ovis sp.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith UA;

XX WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
alpha-S2 casein precursor in the manufacture of a medicament for
alleviating or preventing periodontal disease and an effect of aging in
skin -

PS Claim 8; Page 9; 27pp; English.

CC The invention relates to a composition comprising a peptide or its
derivative. The peptide contains an amino acid sequence from alpha-S2
casein precursor. The peptides stimulate the growth of fibroblasts, and
thus the synthesis and secretion of collagen. The peptides also stimulate
the growth of keratinocytes, which aid in the formation and regeneration
of skin surface. The peptide is useful in the manufacture of a medicament
in the form of a toothpaste or a chewing gum, for alleviating or
preventing periodontal disease and a medicament in the form of a cosmetic
composition for alleviating or preventing an effect of aging,
particularly wrinkling of the skin. The present sequence is sheep
alpha-S2 casein precursor protein.

XX SQ Sequence 223 AA;

Query Match 79.8%; Score 95; DB 23; Length 223;
Best Local Similarity 81.0%; Pred. No. 5.7e-06;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKAMKRWIOPTKTVIPY 21
DB 199 VDOHQKAMKRWIOPTKTVIPY 219

RESULT 15
AAE17475
ID AAE17475 standard; Protein; 223 AA.
XX
AC AAE17475;

XX 22-APR-2002 (first entry)
DT
XX Alpha-S2 casein precursor (alpha-S2-CN) protein #2.

DE Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
KW fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.

XX Unidentified.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith UA;

XX WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
alpha-S2 casein precursor in the manufacture of a medicament for
alleviating or preventing periodontal disease and an effect of aging in
skin -

PS Disclosure; Page 9; 27pp; English.

CC The invention relates to a composition comprising a peptide or its
derivative. The peptide contains an amino acid sequence from alpha-S2
casein precursor. The peptides stimulate the growth of fibroblasts, and
thus the synthesis and secretion of collagen. The peptides also stimulate
the growth of keratinocytes, which aid in the formation and regeneration
of skin surface. The peptide is useful in the manufacture of a medicament
in the form of a toothpaste or a chewing gum, for alleviating or
preventing periodontal disease and a medicament in the form of a cosmetic
composition for alleviating or preventing an effect of aging,
particularly wrinkling of the skin. The present sequence is alpha-S2
casein precursor (alpha-S2-CN) protein.

XX SQ Sequence 223 AA;

Query Match 79.8%; Score 95; DB 23; Length 223;
Best Local Similarity 81.0%; Pred. No. 5.7e-06;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKAMKRWIOPTKTVIPY 21
DB 199 VDOHQKAMKRWIOPTKTVIPY 219

Search completed: July 30, 2003, 16:23:26
Job time : 41.375 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 40.7143 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-3
Perfect score: 135
Sequence: 1 VYQHOKAMKPMWIOPTKTVIPYRVY 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: SP archaea:*
 - 2: SP bacteria:*
 - 3: SP fungi:*
 - 4: SP human:*
 - 5: SP invertebrate:*
 - 6: SP mammal:*
 - 7: SP mhc:*
 - 8: SP organelle:*
 - 9: SP phase:*
 - 10: SP plant:*
 - 11: SP rodent:*
 - 12: SP virus:*
 - 13: SP vertebrate:*
 - 14: SP unclassified:*
 - 15: SP rrvirus:*
 - 16: SP bacteriap:*
 - 17: SP archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	88.1	210	6	062825 bubalus bub
2	111	82.2	223	6	09MYU6 capra hircu
3	102	75.6	223	6	09MYU6 capra hircu
4	102	75.6	223	6	09MYU6 capra hircu
5	53	39.3	935	3	094623 schizoaech
6	52	38.5	80	10	081534 saccharum o
7	52	38.5	85	10	081399 orobanche c
8	52	38.5	340	10	096568 ipomoea pur
9	52	38.5	379	10	004065 perilla fru
10	52	38.5	386	10	023923 digitalis 1
11	52	38.5	388	10	09FS37 torenia hyb
12	52	38.5	389	10	081P19 nierenbergi
13	52	38.5	389	10	043040 pecunia hyb
14	52	38.5	389	10	093XP8 nicotiana t
15	52	38.5	389	10	09MSB2 pecunia hyb
16	52	38.5	390	10	091RB2 scutellaria

17	52	38.5	390	10	048564 scutellaria
18	52	38.5	390	10	022122 scutellaria
19	51	37.8	394	10	093V66 humulus lup
20	51	37.8	399	10	0941W8 humulus lup
21	50	37.0	311	5	019257 caenorbadi
22	50	37.0	693	11	09WV66 mus musculu
23	50	37.0	37.0	3	09USR3 schizoaech
24	49.5	36.7	325	16	09K947 bacillus ha
25	49	36.3	282	5	044074 ascaris suu
26	49	36.3	342	16	08YV06 arabidopsi
27	48.5	35.9	286	10	08W2F2 arabidopsi
28	48	35.6	183	16	09CP51 patenreilla
29	48	35.6	208	16	08F8G2 leptospira
30	48	35.6	225	4	09BTR9 homo sapien
31	48	35.6	328	16	08CP60 strephylloc
32	48	35.6	374	10	081476 bresica na
33	48	35.6	392	10	093YX5 vitis sp. c
34	48	35.6	392	10	09FEK7 vitis labru
35	48	35.6	392	10	081PP4 parthenocis
36	48	35.6	392	10	09FRW2 vitis ripar
37	48	35.6	392	10	081PP2 parthenocis
38	48	35.6	392	10	09S982 vitis. etil
39	48	35.6	392	10	081PP3 clausus rhom
40	48	35.6	392	10	094G58 vitis vinif
41	48	35.6	392	10	0944W8 vitis sp. c
42	48	35.6	392	10	0944W7 vitis sp. c
43	48	35.6	395	10	09FR70 allaria pe
44	48	35.6	395	10	09SLY0 petiolum nu
45	48	35.6	408	10	

ALIGNMENTS

RESULT 1					
ID	062825	PRELIMINARY;	PRT;	210	AA.
AC	062825;				
DT	01-NOV-1998 (TREMBLrel. 07, Created)				
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)				
DT	01-NOV-2001 (TREMBLrel. 17, Last annotation update)				
DE	AS2-casein (Fragment).				
OS	Bubalus bubalis (Domestic water buffalo).				
OC	Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bubalus.				
OX	NCBI_TaxID=89462;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Mammary gland;				
RA	Das P., Jain S., Garg L.C.;				
RT	"Cloning and nucleotide sequence of cDNA encoding AS2-casein in B.				
RT	subalis."				
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AJ005431; CA06534.2; -				
DR	InterPro: IPR001588; Casein.				
DR	Pfam: PF00363; caseins; 2.				
FT	NON_TER				
FT	SEQUENCE				
FT	210 AA; 24700 MW; 05DEF95963F1132C CRC64;				
Query Match	88.1%; Score 119; DB 6; Length 210;				
Best local Similarity	87.5%; Pred. No. 2.4e-10;				
Matches	21; Conservative 1; Mismatches 2; Indels 0; Gaps 0;				
QY	1 VYQHOKAMKPMWIOPTKTVIPYRVY 24				
DB	186 VYQHOKAMKPMWIOPTKTVIPYRVY 209				
RESULT 2					
ID	09MYU6	PRELIMINARY;	PRT;	223	AA.
AC	09MYU6;				

```

DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Alpha s2-casein.
GN CSN182.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Lagomigro R., Pilla F., Mataesino D., Zullo A.;
RT "A new allele of goat alpha s2-casein gene.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ289716; CAB94236.1; -.
DR InterPro: IPR001588; Casein.
DR Pfam: PF00363; caseins; 2.
DR PROSITE: PS00306; CASEIN ALPHA_BETA; 1.
SQ SEQUENCE 223 AA; 26403 MW; 0E1FE83F24DA85E2 CRC64;

Query Match 82.2%; Score 111; DB 6; Length 223;
Best Local Similarity 83.3%; Pred. No. 4.4e-09;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VYOHQKMKPMIOPTKVIPIVRY 24
Db 199 VDHQKMKPMIOPTKVIPIVRY 222

RESULT 3
Q9TTO7 PRELIMINARY; PRT; 223 AA.
AC Q9TTO7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Alpha s2-casein.
GN CSN182.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Veltri C.C., Pilla F.F., Lagomigro R.R.;
RT "A new allele of goat alpha s2-casein.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ249995; CAB59920.1; -.
DR InterPro: IPR001588; Casein.
DR Pfam: PF00363; caseins; 2.
DR PROSITE: PS00306; CASEIN ALPHA_BETA; 1.
SQ SEQUENCE 223 AA; 26433 MW; CE9F4DC8D7688293 CRC64;

Query Match 75.6%; Score 102; DB 6; Length 223;
Best Local Similarity 79.2%; Pred. No. 1.1e-07;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VYOHQKMKPMIOPTKVIPIVRY 24
Db 199 VDHQKMKPMIOPTKVIPIVRY 222

RESULT 4
Q9GK07 PRELIMINARY; PRT; 223 AA.
AC Q9GK07;
DT 01-MAR-2001 (Tremblrel. 15, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Alpha s2-casein.
GN CSN182.

```

```

OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Veltri C., Pilla F., Lagomigro R.;
RT "A new allele of alpha s2-casein.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ297310; CAC21704.2; -.
DR EMBL: AJ297311; CAC21704.2; JOINED.
DR EMBL: AJ242728; CAC21704.2; JOINED.
DR EMBL: AJ297312; CAC21704.2; JOINED.
DR EMBL: AJ297313; CAC21704.2; JOINED.
DR EMBL: AJ297314; CAC21704.2; JOINED.
DR EMBL: AJ242527; CAC21704.2; JOINED.
DR EMBL: AJ297315; CAC21704.2; JOINED.
DR EMBL: AJ242526; CAC21704.2; JOINED.
DR EMBL: AJ242528; CAC21704.2; JOINED.
DR EMBL: AJ242533; CAC21704.2; JOINED.
DR InterPro: IPR001588; Casein.
DR Pfam: PF00363; caseins; 2.
DR PROSITE: PS00306; CASEIN ALPHA_BETA; 1.
SQ SEQUENCE 223 AA; 26432 MW; CE9765E8D7688C9D CRC64;

Query Match 75.6%; Score 102; DB 6; Length 223;
Best Local Similarity 79.2%; Pred. No. 1.1e-07;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VYOHQKMKPMIOPTKVIPIVRY 24
Db 199 VDHQKMKPMIOPTKVIPIVRY 222

RESULT 5
O94623 PRELIMINARY; PRT; 935 AA.
AC O94623;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical 106.5 kDa protein C1347.01C in chromosome II.
GN SPBC1347.01C OR SPBC215.16C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RC Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE OF 731-935 FROM N.A.
RC STRAIN=972;
RA Lyne M., Rajandream M.A., Barrell B.G., Rieger M.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: TO C.ELEGANS ZK675.2.
CC -1- SIMILARITY: BELONGS TO THE UMC (DNA REPAIR) FAMILY.
CC -1- SIMILARITY: CONTAINS 1 BRCT DOMAIN.
DR EMBL: AL035548; CAB37432.1; -.
DR EMBL: AL033534; CAA22130.1; -.
DR HSP: P96022; 1M4.
DR GeneDB SPombe: SPBC1347.01c; -.
DR InterPro: IPR001357; BRCT.
DR InterPro: IPR001126; UMC-like.
DR Pfam: PF00553; BRCT; 1.
DR Pfam: PF00817; IMS; 1.
DR SMART: SM00292; BRCT; 1.

```


RA Gong Z., Yamazaki M., Sugiyama M., Tanaka Y., Saito K.;
RT "Cloning and molecular analysis of structural genes involved in
RT anthocyanin biosynthesis and expressed in a forma-specific manner in
RT *Perilla frutescens*.";

RU Plant Mol. Biol. 35:915-927(1997).
DR EMBL; AB002582; BAA19548.1; -.
DR HSSP; P30074; 1D6F.
DR InterPro; IPR001099; N-C_synthase.
DR Pfam; PF00195; Chal_stil_synth; 1.
DR Pfam; PF02797; Chal_stil_synthC; 1.
DR Prodom; PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
SQ SEQUENCE 379 AA; 41380 MW; 5E149C55530BD106 CRC64;

QY Query Match 38.5%; Score 52; DB 10; Length 379;
Best Local Similarity 64.3%; Pred. No. 8.9;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 111 OKAIKEMGQPKSKI 124

RESULT 10

OC Q23923 PRELIMINARY; PRT; 386 AA.
ID Q23923;
AC Q23923;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Chalcone synthase (Foxglove).
OS Digitalis lanata (Foxglove).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Lamiales; Antirrhinaceae; Digitalidae; Digitalis.
NCBI_TaxID=19450;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RA Thoeninger C., Lindemann P., Luckner M.;
RT "Expression of phenylalanine ammonia lyase, chalcone synthase and
RT phytochrome in somatic embryogenesis of *Digitalis lanata*.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002526; CAA05512.1; -.
DR HSSP; P30074; 1D6F.
DR InterPro; IPR001099; N-C_synthase.
DR Pfam; PF00195; Chal_stil_synth; 1.
DR Pfam; PF02797; Chal_stil_synthC; 1.
DR Prodom; PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
FT NON TER 1 1
SQ SEQUENCE 386 AA; 42190 MW; 0A8B8F6EDA82F8F CRC64;

QY Query Match 38.5%; Score 52; DB 10; Length 386;
Best Local Similarity 64.3%; Pred. No. 9.1;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 106 OKAIKEMGQPKSKI 119

RESULT 11

OC Q9FS37 PRELIMINARY; PRT; 368 AA.
ID Q9FS37;
AC Q9FS37;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Chalcone synthase.
OS *Torenia hybrida*.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Lamiales; Lamiales incertae sedis; Torenieae;
OC *Torenia*.
NCBI_TaxID=75807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Summerwave Blue; TISSUE=Petal;
RA Suzuki K., Xue H., Tanaka Y., Fukui Y., Fukuchi-Mitani M.,
RA Murakami Y., Katsumoto Y., Tsuda S., Kusumi T.;
RT "Flower color modifications of *Torenia hybrida* by cosuppression of
RT anthocyanin biosynthesis genes.";

RU Mol. Breed. 6:239-246(2000).
DR EMBL; AB012923; BAB20074.1; -.
DR HSSP; P30074; 1D6F.
DR InterPro; IPR001099; N-C_synthase.
DR Pfam; PF00195; Chal_stil_synth; 1.
DR Pfam; PF02797; Chal_stil_synthC; 1.
DR Prodom; PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
SQ SEQUENCE 388 AA; 42402 MW; BD3A78B6C4A6B0F CRC64;

QY Query Match 38.5%; Score 52; DB 10; Length 388;
Best Local Similarity 64.3%; Pred. No. 9.2;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 111 OKAIKEMGQPKSKI 124

RESULT 12

OC Q8LP19 PRELIMINARY; PRT; 389 AA.
ID Q8LP19;
AC Q8LP19;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Chalcone synthase.
OS *Nierembergia* sp. NB17.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nierembergia.
NCBI_TaxID=184877;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NB17;
RA Ueyama Y., Katsumoto Y., Fukui Y., Ohkawa H., Kusumi T., Tanaka Y.;
RT "Flower color modification of *Nierembergia* sp. by engineering
RT flavonoid biosynthetic pathway.";

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB078515; BAC10998.1; -.
DR InterPro; IPR001099; N-C_synthase.
DR Pfam; PF00195; Chal_stil_synth; 1.
DR Pfam; PF02797; Chal_stil_synthC; 1.
DR Prodom; PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
SQ SEQUENCE 389 AA; 42616 MW; 62DA0C7F407729E7 CRC64;

QY Query Match 38.5%; Score 52; DB 10; Length 389;
Best Local Similarity 64.3%; Pred. No. 9.2;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 111 OKAIKEMGQPKSKI 124

RESULT 13

OC Q43040 PRELIMINARY; PRT; 389 AA.
ID Q43040;
AC Q43040;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 GN CHS-A.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 NCBI_TaxId=4102;
 RX MEDLINE=96156377; PubMed=8562849;
 RA Shao L., Li Y., Pan A., Cheng Z., Chen M.;
 RT "Molecular cloning, sequencing, and expression in *Escherichia coli* of
 the chalcone synthase gene."
 RL Chin. J. Biotechnol. 11:131-135(1995).
 DR EMBL; S80857; AAB36038.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 389 AA; 42642 MW; A931CF5B255A0A20 CRC64;

Query Match 38.5%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 9.2;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
 DB 111 OKAIKEWQPKSKI 124

RESULT 14
 Q93XP8 PRELIMINARY; PRT; 389 AA.
 ID Q93XP8
 AC Q93XP8;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Chalcone synthase.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 NCBI_TaxId=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Flower;
 RA Hu H., Sung H., Su J.;
 RT "Cloning and Expression of a Tobacco Chalcone Synthase Gene."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF311783; AAK49457.1; -.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 389 AA; 42563 MW; 96C3D426786E8D44 CRC64;

Query Match 38.5%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 9.2;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
 DB 111 OKAIKEWQPKSKI 124

RESULT 15
 Q9MSB2 PRELIMINARY; PRT; 389 AA.
 ID Q9MSB2
 AC Q9MSB2;

DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Chalcone synthase (EC 2.3.1.74).
 GN CHS.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 NCBI_TaxId=4102;
 RX MEDLINE=96156377; PubMed=8562849;
 RA Shao L., Li Y., Pan A., Cheng Z., Chen M.;
 RT "Molecular cloning, sequencing, and expression in *Escherichia coli* of
 the chalcone synthase gene."
 RL Chin. J. Biotechnol. 11:131-135(1995).
 DR EMBL; S80857; AAB36038.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_stil_synth; 1.
 DR Pfam; PF02797; Chal_stil_synth; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
 KW Acyltransferase; Transferase.
 SQ SEQUENCE 389 AA; 42582 MW; 66135700259D7F0C CRC64;

Query Match 38.5%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 9.2;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
 DB 111 OKAIKEWQPKSKI 124

Search completed: July 30, 2003, 16:29:22
 Job time : 40.7143 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 30, 2003, 16:10:10 ; Search time 7.85714 Seconds

(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-3

Perfect score: 135
Sequence: 1 VYQHOKAMKPMWIOPTKIVPYRY 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135	100.0	222	1	CAS2_BOVIN
2	111	82.2	223	1	CAS2_CAPI
3	111	82.2	223	1	CAS2_SHEEP
4	70.5	52.2	235	1	CAS2_PIG
5	67	49.6	182	1	CAS3_RABIT
6	55	40.7	193	1	CAS2_CAMDR
7	52	38.5	389	1	CHS1_LYCES
8	52	38.5	389	1	CHS2_LYCES
9	52	38.5	389	1	CHS2_SOLTU
10	52	38.5	389	1	CHS2_PETHY
11	52	38.5	389	1	CHS2_SOLTU
12	52	38.5	389	1	CHS2_SOLTU
13	52	38.5	389	1	CHS2_IPOPU
14	52	38.5	389	1	CHS2_PETHY
15	52	38.5	389	1	CHS2_PETHY
16	52	38.5	389	1	CHS2_PETHY
17	52	38.5	389	1	CHS2_PETHY
18	52	38.5	389	1	CHS2_PETHY
19	52	38.5	389	1	CHS2_PETHY
20	52	38.5	389	1	CHS2_PETHY
21	52	38.5	389	1	CHS2_PETHY
22	52	38.5	389	1	CHS2_PETHY
23	52	38.5	389	1	CHS2_PETHY
24	52	38.5	389	1	CHS2_PETHY
25	50	37.0	221	1	PCIL_MOUSE
26	49.5	36.7	207	1	VEGB_MOUSE
27	49	36.3	392	1	CHS1_SECCB
28	49	36.3	392	1	CHS1_SECCB
29	49	36.3	392	1	CHS1_SECCB
30	49	36.3	392	1	CHS1_SECCB
31	49	36.3	392	1	CHS1_SECCB
32	48	35.6	389	1	CHS1_VITVI
33	48	35.6	392	1	THS1_VITVI

34	48	35.6	392	1	THS2_VITVI
35	48	35.6	5171	1	BPEA_HUMAN
36	47	34.8	285	1	CHS4_MEDSA
37	47	34.8	311	1	CHS4_TRISU
38	47	34.8	388	1	CHS1_SOYBN
39	47	34.8	388	1	CHS2_SOYBN
40	47	34.8	388	1	CHS3_SOYBN
41	47	34.8	388	1	CHS5_SOYBN
42	47	34.8	388	1	CHS6_SOYBN
43	47	34.8	388	1	CHS8_IPOBA
44	47	34.8	388	1	CHSD_PETHY
45	47	34.8	388	1	CHS_VIGUN

ALIGNMENTS

RESULT 1	CAS2_BOVIN	STANDARD;	PRT;	222 AA.
AC	P02663; Q9TR51;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-JUL-1989 (Rel. 11, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Alpha-52 casein precursor (Contains: Casocidin-1).			
GN	CSN1S2.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=6818899; PubMed=2833659;			
RA	Stewart A.F., Bomsing J., Beattie C.W., Shah F., Willis I.M.,			
RA	Mackinley A.G.;			
RT	"Complete nucleotide sequences of bovine alpha S2- and beta-casein			
RT	cDNAs: comparisons with related sequences in other species.";			
RL	Mol. Biol. Evol. 4:231-241(1987).			
RN	[2]			
RP	SEQUENCE OF 16-222 (A ALLELE).			
RC	TISUE=Milk;			
RX	MEDLINE=77185633; PubMed=862906;			
RA	Brignon G., Ribadeau-Dumas B., Mercier J.-C., Pelissier J.-P.,			
RA	Das B.C.;			
RT	"Complete amino acid sequence of bovine alphaS2-casein.";			
RL	FEBS Lett. 76:274-279(1977).			
RN	[3]			
RP	PARTIAL SEQUENCE (D ALLELE).			
RC	TISUE=Milk;			
RX	MEDLINE=79239837; PubMed=469044;			
RA	Grosclaude F., Joudrier P., Mahe M.-F.,			
RT	"A genetic and biochemical analysis of a polymorphism of bovine alpha			
RT	S2-casein.";			
RL	J. Dairy Res. 46:211-213(1979).			
RN	[4]			
RP	SEQUENCE OF 165-203, AND CHARACTERIZATION OF CASOCIDIN.			
RC	TISUE=Milk;			
RX	MEDLINE=96000204; PubMed=7556666;			
RA	Zucht H.-D., Ralda M., Adernann K., Meagert H.-J., Forssmann W.-G.;			
RT	"Casocidin-I: a casein-alpha s2 derived peptide exhibits antibacterial			
RT	activity.";			
RL	FEBS Lett. 372:185-188(1995).			
CC	-1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT			
CC	CALCIUM PHOSPHATE.			
CC	-1- FUNCTION: CASOCIDIN-1 INHIBITS THE GROWTH OF ESCHERICHIA COLI AND			
CC	STAPHYLOCOCCUS CARNEUS.			
CC	-1- SUBCELLULAR LOCATION: Extracellular.			
CC	-1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.			
CC	-1- MASS SPECTROMETRY: MW=4870; METHOD=Electrospray; RANGE=165-203.			
CC	-1- POLYMORPHISM: AT LEAST TWO ALLELES EXIST. THE SEQUENCE OF THE A			
CC	ALLELE IS SHOWN HERE. THE D ALLELE SEQUENCE DIFFERS FROM THAT			
CC	SHOWN IN HAVING A DELETION OF NINE RESIDUES, WHICH MAY BE 49-58,			

CC 50-59, OR 51-60.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC DATABASE: NAME=Protein Spotlight;
 CC NOTE=Issue 16 of November 2001;
 CC WWW=http://www.expasy.org/spotlight/articles/sp16016.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M16644; AAA30479.1; -
 CC PIR; JQ2008; KABOS2.
 CC InterPro: IPR001588; Casein.
 CC Pfam; PF00363; caseins; 2.
 CC DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 CC KW Milk; Phosphorylation; Signal; Repeat; Antibiotic.
 CC FT SIGNAL 1 15
 CC FT CHAIN 16 222 ALPHA-S2 CASEIN.
 CC FT PEPTIDE 165 203 CASOCIDIN-I.
 CC FT REPEAT 158 222
 CC FT REPEAT 158 222
 CC FT MOD_RES 23 23 PHOSPHORYLATION.
 CC FT MOD_RES 24 24 PHOSPHORYLATION.
 CC FT MOD_RES 25 25 PHOSPHORYLATION.
 CC FT MOD_RES 31 31 PHOSPHORYLATION.
 CC FT MOD_RES 71 71 PHOSPHORYLATION.
 CC FT MOD_RES 72 72 PHOSPHORYLATION.
 CC FT MOD_RES 73 73 PHOSPHORYLATION.
 CC FT MOD_RES 76 76 PHOSPHORYLATION.
 CC FT MOD_RES 144 144 PHOSPHORYLATION.
 CC FT MOD_RES 146 146 PHOSPHORYLATION.
 CC FT MOD_RES 102 102 Q -> E (IN REP. 2).
 CC FT SIGNAL 222 222 AA; 81E7408AF1C12FC CRC64;
 CC SQ SEQUENCE
 CC Query Match 100.0%; Score 135; DB 1; Length 222;
 CC Best Local Similarity 100.0%; Pred. No. 9.6e-13;
 CC Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC QY 1 VYOHQAMKPMIOPKTKVPIPVRY 24
 CC DB 198 VYOHQAMKPMIOPKTKVPIPVRY 221
 CC RESULT 2
 CC CAS2_CAPHI STANDARD; PRT; 223 AA.
 CC ID CAS2_CAPHI
 CC AC P33049;
 CC DT 01-OCT-1993 (Rel. 27, Created)
 CC DT 01-OCT-1993 (Rel. 27, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Alpha-S2 casein precursor (Alpha-S2-CN).
 CC GN CSN152.
 CC OS Capra hircus (Goat).
 CC OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC OC Bovidae; Caprinae; Capra.
 CC OC NCBI_TaxID=9925;
 CC OX [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=93216130; PubMed=8462880;
 CC RA Bouniol C.; Brignon G.; Mahe M.-F.; Printz C.;
 CC RT "Sequence of the goat alpha s2-casein-encoding cDNA";
 CC RL Gene 125:235-236(1993).
 CC RN [2]
 CC RP SEQUENCE FROM N.A.
 CC RA Bouniol C.; Brignon G.; Mahe M.-F.; Printz C.;
 CC RT "Characterization of goat allelic alpha-s2-caseins A and B: further
 CC evidence of the phosphorylation code of caseins";
 CC RL Protein Seq. Data Anal. 5:213-218(1993).
 CC RN [3]

RN [3]
 RN SEQUENCE FROM N.A. (VARIANT C).
 RN MEDLINE=95030556; PubMed=7943951;
 RN Bouniol C.; Brignon G.; Mahe M.-F.; Printz C.;
 RN RT "Biochemical and genetic analysis of variant C of caprine alpha s2-
 RN casein (Capra hircus)".
 RN RL Anim. Genet. 25:173-177(1994).
 CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
 CC CALCIUM PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- POLYMORPHISM: THREE ALLELES OF ALPHA-S2 CASEIN ARE KNOWN: A, B AND
 CC C. THE FREQUENCIES OF THE ALLELES IS ESTIMATED TO BE 0.85, 0.04
 CC AND 0.11 IN THE FRENCH DAIRY BREEDS 'ALPINE' AND 'SAANEN'.
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X65160; CAA46278.1; -
 CC EMBL; S74171; AAB32166.1; -
 CC PIR; I46995; I46995.
 CC PIR; S33881; JN0547.
 CC InterPro: IPR001588; Casein.
 CC Pfam; PF00363; caseins; 2.
 CC DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
 CC KW Milk; Phosphorylation; Signal; Repeat; Polymorphism.
 CC FT SIGNAL 1 15
 CC FT CHAIN 16 223 ALPHA-S2 CASEIN.
 CC FT REPEAT 159 223
 CC FT REPEAT 159 223
 CC FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 77 77 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 145 145 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 147 147 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 147 147 PHOSPHORYLATION (BY SIMILARITY).
 CC FT VARIANT 79 79 E -> K (IN VARIANT B).
 CC FT VARIANT 182 182 K -> I (IN VARIANT C).
 CC FT SIGNAL 223 223 AA; 187DEFA2FD688291 CRC64;
 CC SQ SEQUENCE
 CC Query Match 82.2%; Score 111; DB 1; Length 223;
 CC Best Local Similarity 83.3%; Pred. No. 3.2e-09;
 CC Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 CC QY 1 VYOHQAMKPMIOPKTKVPIPVRY 24
 CC DB 199 VYOHQAMKPMIOPKTKVPIPVRY 222
 CC RESULT 3
 CC CAS2_SHEEP STANDARD; PRT; 223 AA.
 CC ID CAS2_SHEEP
 CC AC P04654;
 CC DT 13-AUG-1987 (Rel. 05, Created)
 CC DT 13-AUG-1987 (Rel. 05, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Alpha-S2 casein precursor.
 CC GN CSN152.
 CC OS Ovis aries (Sheep).
 CC OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC OC Bovidae; Caprinae; Ovis.
 CC OC NCBI_TaxID=9940;
 CC OX [1]


```

RP SEQUENCE FROM N.A.
RX MEDLINE=6104467; PubMed=3002499;
RA Boissard M., Petrisant G.;
RT "Complete sequence of ovine alpha s2-casein messenger RNA.";
RL Biochimie 67:1043-1051(1985)
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03238; CAA26983.1; -.
DR PIR; A25070; KASHS2.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
KW Milk; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 15
FT CHAIN 16 223 ALPHA-S2 CASEIN.
FT REPEAT 77 141
FT REPEAT 159 223
FT MOD_RES 23 223 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 24 224 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 25 225 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 72 72 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 73 73 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 74 74 PHOSPHORYLATION (POTENTIAL).
FT VARIANT 64 64 D -> N.
SQ SEQUENCE 223 AA; 26332 MW; 67212935E27426D7 CRC64;

Query Match 82.2%; Score 111; DB 1; Length 223;
Best Local Similarity 83.3%; Pred. No. 3.2e-09;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHQKAMKRWIQPKTKVPIPVYRY 24
DB 199 VDHQKAMKRWIQPKTKNAIPVYRY 222

RESULT 4
CAS2_PIG STANDARD; PRT; 235 AA.
ID CAS2_PIG
AC P39036;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-S2 casein precursor.
GN CSN1S2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sub.
OC NCBI_TaxId=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92367960; PubMed=1503276;
RA Alexander L.J., Das Gupta N.A., Beattie C.W.;
RT "The sequence of porcine alpha s2-casein cDNA.";
RL Anim. Genet. 23:365-367(1992).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54975; CAA38719.1; -.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
KW Milk; Phosphorylation; Signal.
FT SIGNAL 1 15 BY SIMILARITY.
FT CHAIN 16 235 ALPHA-S2 CASEIN.
FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 70 70 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 71 71 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 235 AA; 27570 MW; C903B760D184C14C CRC64;

Query Match 52.2%; Score 70.5; DB 1; Length 235;
Best Local Similarity 55.6%; Pred. No. 0.003;
Matches 15; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

QY 1 VYOHQKAMKRWIQPKTKVPIPVYRY 24
DB 208 VHQKQAMKRWIHKINSYQIIPNRY 234

RESULT 5
CAS3_RABIT STANDARD; PRT; 182 AA.
ID CAS3_RABIT
AC P50419;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Alpha-S2B casein precursor.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OC NCBI_TaxId=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=New Zealand white; TISSUE=Mammary gland;
RX MEDLINE=94107245; PubMed=8280077;
RA Dawson S.P., Wilde C.J., Tighe P.J., Mayer R.J.;
RT "Characterization of two novel casein transcripts in rabbit mammary
RT gland.";
RL Biochem. J. 296:777-784(1993).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76909; CAA54231.1; -.
DR PIR; S39776; S39776.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.

```

```

DR PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
KW Milk; Phosphorylation; signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 162 AA; 21597 MW; F2432C2DA4550FC CRC64;
OY
Query Match 49.6%; Score 67; DB 1; Length 182;
Best Local Similarity 43.5%; Pred. No. 0.0077;
Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;
Db 83 LYQYPTVMDPWTBETKAIPIFR 105
OY 1 VYOHQKAKMPWIOPTKVIPIYVR 23
::: ||| : ||| : ||| :
AC CAS2_CAMDR STANDARD; PRT; 193 AA.
AC 097944;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Alpha-S2 casein precursor.
GN CSN1S2.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Somali; TISSUE=udder;
RX MEDLINE=98291310; PubMed=9627940;
RA Kappeler S., Farah Z., Puhaz Z.;
RT "Sequence analysis of Camelus dromedarius milk caseins.";
RL J. Dairy Res. 65:209-222(1998).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AJ012622; CA110078.1; -.
DR InterPro; IPR001588; Casein.
DR Pfam; PF00363; caseins; 2.
DR PROSITE; PS00306; CASEIN_ALPHA_BETA; FALSE_NEG.
KW Milk; Phosphorylation; signal.
FT SIGNAL
FT CHAIN
FT SEQUENCE 162 AA; 22964 MW; 2843256F9FDD2ED13 CRC64;
SQ
OY
Query Match 40.7%; Score 55; DB 1; Length 193;
Best Local Similarity 40.9%; Pred. No. 0.47;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
Db 93 LHQGIYVNNPMDGKTRAVPI 114
OY 1 VYOHQKAKMPWIOPTKVIPIYVR 22
::: ||| : ||| : ||| :
AC CHS1_LYCES STANDARD; PRT; 389 AA.
AC P23418;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)

```

```

DE      15-JUN-1999 (Rel. 38, last annotation update)
DT      Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1).
GN      CHS1.
OC      Lycopersicon esculentum (Tomato).
OS      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
RX      Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
OC      Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX      NCBI_TaxID=4081;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Cotyledon, Hypocotyl, and Leaf;
RX      MEDLINE=91117196; PubMed=1980524;
RA      O'Neill S.D., Tong Y., Speerlein B., Forkmann G., Yoder J.I.;
RT      "Molecular genetic analysis of chalcone synthase in Lycopersicon
RT      esculentum and an anthocyanin-deficient mutant.";
RJ      Mol. Genet. 224:279-288(1990) .
CC      -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC      tetrahydrochalcone (also termed naringenin-chalcone or chalcone)
CC      which can under specific conditions spontaneously isomerize into
CC      naringenin.
CC      -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC      naringenin-chalcone + 3 CO(2).
CC      -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC      FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC      OF WHICH ARE BRIGHLY COLORED.
CC      -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-eb.ch/announce/
CC      or send an email to license@isb-eb.ch).
-----
CC      EMBL; X55194; CAAB3980.1; -.
DR      InterPro: IPR001099; N-C_synthase.
DR      Pfam; PF00195; Chal_stil_syntc; 1.
DR      Pfam; PF02797; Chal_stil_syntc; 1.
DR      ProDom; PD000453; N-C_synthase; 1.
DR      PROSITE; PS000441; CHALCONE_SYNTH; 1.
KW      Flavonoid biosynthesis; Transferase; Acyltransferase;
KW      Multigene family.
FT      ACT_SITE 164..164 BY SIMILARITY
SQ      SEQUENCE 389 AA; 42552 MW; 553D65B5EA6A8B CRC64;
Query Match 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 2.6;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Oy      5 QKANKPMIOPKTV 18
Db      111 QKAIKWGPGRSKI 124
RESULT 8
CHS2_LYCES STANDARD; PRT; 389 AA.
AC      P23419;
DT      01-NOV-1991 (Rel. 20, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
GN      CHS2.
OS      Lycopersicon esculentum (Tomato).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
OC      Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX      NCBI_TaxID=4081;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Cotyledon, Hypocotyl, and Leaf;
RX      MEDLINE=91117196; PubMed=1980524;

```

RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
 RT "Molecular genetic analysis of chalcone synthase in *Lycopersicon*
 RL esculentum and an anthocyanin-deficient mutant.",
 CC Mol. Gen. Genet. 224:279-286(1990).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X55195; CA38981.1; -.
 DR InterPro: IPR001099; N-C synthase.
 DR Pfam: PF00195; Chal_scl1_synt; 1.
 DR Pfam: PF02797; Chal_scl1_synt; 1.
 DR ProDom: PD000453; N-C synthase; 1.
 DR PROSITE: PS000441; CHALCONE_SYNTH; 1.
 KM Flavonoid biosynthesis; Transferase; Acyltransferase;
 KW Multigene family.
 FT ACT SITE 164
 SQ SEQUENCE 389 AA; 42730 MW; F92B46B3B5FC32F CRC64;

Query Match 38.5%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 2.6;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
 Db 111 OKAIKMGQPKSKI 124

RESULT 9
 CHS2 SOLTU STANDARD; PRT; 389 AA.
 AC 043188;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
 GN CHS2.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Red Pontiac;
 RX MEDLINE=97141614; PubMed=8987872;
 RA Jeon J.-H., Kim H.S., Choi K.H., Joung Y.H., Joung H., Byun S.-M.;
 RT "Cloning and characterization of one member of the chalcone synthase
 RT gene family from *Solanum tuberosum* L.",
 RL Biosci. Biotechnol. Biochem. 60:1907-1910(1996).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY

CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U47738; AAB05239.1; -.
 DR PIR: JC5136; JC5136.
 DR InterPro: IPR001099; N-C synthase.
 DR Pfam: PF00195; Chal_scl1_synt; 1.
 DR Pfam: PF02797; Chal_scl1_synt; 1.
 DR ProDom: PD000453; N-C synthase; 1.
 DR PROSITE: PS000441; CHALCONE_SYNTH; 1.
 KM Flavonoid biosynthesis; Transferase; Acyltransferase;
 KW Multigene family.
 FT ACT SITE 164
 SQ SEQUENCE 389 AA; 42476 MW; 41618F94495603 CRC64;

Query Match 38.5%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 2.6;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
 Db 111 OKAIKMGQPKSKI 124

RESULT 10
 CHS4 PERTHY STANDARD; PRT; 389 AA.
 AC P08894;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase A (EC 2.3.1.74) (Naringenin-chalcone synthase A).
 GN CHS4.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 OX NCBI_TaxID=4102;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Violet 30;
 RX MEDLINE=86286540; PubMed=3016642;
 RA Koes R.E., Spelt C.E., Reif H.J., van den Elzen P.J.M., Veltkamp E.,
 RA Mol J.N.M.;
 RT "Floral tissue of *Petunia hybrida* (V30) expresses only one member of
 RT the chalcone synthase multigene family.",
 RL Nucleic Acids Res. 14:5229-5239(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Violet 30; TISSUE=Leaf;
 RX MEDLINE=90034197; PubMed=2806915;
 RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;
 RT "Cloning and molecular characterization of the chalcone synthase
 RT multigene family of *Petunia hybrida*.",
 RL Gene 81:245-257(1989).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- TISSUE SPECIFICITY: MAJOR EXPRESSED MEMBER OF THE GENE FAMILY IN

VARIOUS FLORAL TISSUES AND IN SEEDLINGS TREATED WITH UV LIGHT. IT IS RELATIVELY LOW EXPRESSED IN TISSUE CULTURE MATERIAL.
-1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; X04080; CAA27718.1; -
CC EMBL; X14591; CAA32731.1; -
CC PIR; A23643; SYPUJN.
CC PIR; US0308; SYPUJA.
CC InterPro: IPR001099; N-C_synthase.
CC Pfam; PF00195; Chal_stil_synth; 1.
CC Pfam; PF02797; Chal_stil_synthC; 1.
CC ProDom; PD000453; N-C_synthase; 1.
CC PROSITE; PS00441; CHALCONE SYNTH; 1.
CC Flavonoid biosynthesis; Transferase; Acyltransferase;
CC Multigene family.
CC ACT SITE 164 BY SIMILARITY.
CC FT CONFLICT 128 V -> F (IN REF. 2).
CC SQ SEQUENCE 389 AA; 42525 MW; 0221B8DA1E96EAF6 CRC64;

Query Match 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 2.6;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QKAKMPWIOPTKV 18
Db 111 QKAKMPWIOPTKV 124

RESULT 11
CHSB SOLUTV STANDARD; PRT; 389 AA.
AC 041436;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase 1A (EC 2.3.1.74) (Naringenin-chalcone synthase 1A).
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_Taxid=4113;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Red Pontiac;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Young Y.-H., Young H., Byun S.-M.;
RT "Characterization of two members of the chalcone synthase gene family
RT from Solanum tuberosum L.";
RL (in) Plant Gene Register PGR96-027.
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U47740; AAB67735.1; -
CC InterPro: IPR001099; N-C_synthase.
CC Pfam; PF00195; Chal_stil_synth; 1.
CC Pfam; PF02797; Chal_stil_synthC; 1.
CC ProDom; PD000453; N-C_synthase; 1.
CC PROSITE; PS00441; CHALCONE SYNTH; 1.
CC Flavonoid biosynthesis; Transferase; Acyltransferase;
CC Multigene family.
CC ACT SITE 164 BY SIMILARITY.
CC FT CONFLICT 164
CC SQ SEQUENCE 389 AA; 42548 MW; E7334A1A34D8D1CC CRC64;

entites requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; U47739; AAB67734.1; -
CC InterPro: IPR001099; N-C_synthase.
CC Pfam; PF00195; Chal_stil_synth; 1.
CC Pfam; PF02797; Chal_stil_synthC; 1.
CC ProDom; PD000453; N-C_synthase; 1.
CC PROSITE; PS00441; CHALCONE SYNTH; 1.
CC Flavonoid biosynthesis; Transferase; Acyltransferase;
CC Multigene family.
CC ACT SITE 164 BY SIMILARITY.
CC FT CONFLICT 164
CC SQ SEQUENCE 389 AA; 42562 MW; B1B1D9C6B9170A34 CRC64;

Query Match 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 2.6;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 QKAKMPWIOPTKV 18
Db 111 QKAKMPWIOPTKV 124

RESULT 12
CHSB SOLUTV STANDARD; PRT; 389 AA.
AC 041463;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase 1B (EC 2.3.1.74) (Naringenin-chalcone synthase 1B).
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_Taxid=4113;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Red Pontiac;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Young Y.-H., Young H., Byun S.-M.;
RT "Characterization of two members of the chalcone synthase gene family
RT from Solanum tuberosum L.";
RL (in) Plant Gene Register PGR96-027.
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U47740; AAB67735.1; -
CC InterPro: IPR001099; N-C_synthase.
CC Pfam; PF00195; Chal_stil_synth; 1.
CC Pfam; PF02797; Chal_stil_synthC; 1.
CC ProDom; PD000453; N-C_synthase; 1.
CC PROSITE; PS00441; CHALCONE SYNTH; 1.
CC Flavonoid biosynthesis; Transferase; Acyltransferase;
CC Multigene family.
CC ACT SITE 164 BY SIMILARITY.
CC FT CONFLICT 164
CC SQ SEQUENCE 389 AA; 42548 MW; E7334A1A34D8D1CC CRC64;


```

OS  Petunia hybrida (Petunia).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
OC  Asteridae; lamiales; Solanales; Solanaceae; Petunia.
OX  NCBI_TaxID=4102.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=cv. Violet 30; TISSUE=leaf;
RX  MEDLINE=90034197; PubMed=2806915;
RA  Koes R.E., Spelt C.E., van den Bizen P.J.M., Mol J.N.M.;
RT  "Cloning and molecular characterization of the chalcone synthase
RL  multigene family of Petunia hybrida.";
    Gene 81:245-257(1989).
RN  [2]
RP  SEQUENCE OF 71-389 FROM N.A.
RC  STRAIN=cv. White 137; TISSUE=anther;
RA  van Tunen A.J.;
RL  Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
CC  tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
CC  which can under specific conditions spontaneously isomerize into
CC  naringenin.
CC  -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC  naringenin-chalcone + 3 CO(2).
CC  -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC  FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC  OF WHICH ARE BRIGHTLY COLORED.
CC  -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X14597; CAA32737.1; -.
DR  EMBL; X14599; CAA32739.1; -.
DR  PIR; D72821; STPJGJ.
DR  PIR; S18136; S18136.
DR  InterPro: IPR001099; N-C_synthase.
DR  Pfam; PF00195; Chal_stil_synth; 1.
DR  Pfam; PF02797; Chal_stil_synth; 1.
DR  ProDom; PD000453; N-C_synthase; 1.
DR  PROSITE; PS00441; CHALCONE SYNTH; 1.
KM  Flavonoid biosynthesis; Transferase; Acyltransferase;
KM  Multigene family.
FT  ACT_SITE 164 164 BY SIMILARITY.
FT  CONFLICT 75 75 E -> V (IN REF. 2).
SQ  SEQUENCE 389 AA; 42558 MW; F2B3CDD82E6FDE7D CRC64;

```

```

Query March 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 2.6;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 5 OKAMKPMIOPTKV 18
    |||:| |||:|
DB 111 OKAIKMGQPKSKI 124

```

Search completed: July 30, 2003, 16:24:28
 Job time : 7.85714 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 15.2857 Seconds

(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-3

Perfect score: 135
Sequence: 1 VYQHOKAMKRWIQPKTKIPYVRY 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	135	100.0	222	1 KABS02	alpha-s2-casein pr
2	111	82.2	223	1 KASHS2	alpha-s2-casein pr
3	111	82.2	223	2 JN0547	alpha-s2-casein pr
4	111	82.2	223	2 I46995	alpha-s2-casein C
5	70.5	52.2	235	2 A48383	alpha-s2-casein -
6	67	49.6	182	2 S39776	alpha-s2-casein fo
7	53	39.3	935	2 T39389	probable DNA damag
8	52	38.5	319	2 S18136	naringenin-chalcon
9	52	38.5	340	2 T10957	naringenin-chalcon
10	52	38.5	389	1 SYPOCU	naringenin-chalcon
11	52	38.5	389	1 SYPOCN	naringenin-chalcon
12	52	38.5	389	1 SYPOCA	naringenin-chalcon
13	52	38.5	389	2 JCS136	naringenin-chalcon
14	52	38.5	389	2 T07799	naringenin-chalcon
15	52	38.5	390	1 SYSKCD	naringenin-chalcon
16	51	37.8	180	2 S39775	alpha-s2-casein fo
17	50	37.0	311	2 T15997	hypothetical prote
18	49.5	36.7	1877	2 T40550	probable phosphat
19	49.5	36.7	325	2 C84000	hypothetical prote
20	49.5	36.7	564	2 T11548	probable zinc meta
21	49	36.3	188	2 JCS680	vascular endotheli
22	49	36.3	207	2 JCS679	vascular endotheli
23	49	36.3	342	2 AD2032	hypothetical prote
24	49	36.3	398	2 S16275	naringenin-chalcon
25	49	36.3	400	1 SYZMCC	naringenin-chalcon
26	48	35.6	392	2 S16206	stilbene synthase
27	48	35.6	392	2 S11044	stilbene synthase
28	48	35.6	392	2 S5313	stilbene synthase
29	48	35.6	392	2 S5314	stilbene synthase

30	48	35.6	418	2 F72370	phoH-related prote
31	47	34.8	53	2 S41957	naringenin-chalcon
32	47	34.8	128	2 S33611	naringenin-chalcon
33	47	34.8	158	2 B66169	hypothetical prote
34	47	34.8	285	2 S44369	naringenin-chalcon
35	47	34.8	331	2 JQ1071	naringenin-chalcon
36	47	34.8	370	2 S44368	naringenin-chalcon
37	47	34.8	383	2 S35165	naringenin-chalcon
38	47	34.8	388	1 SYGYCN	naringenin-chalcon
39	47	34.8	388	1 SYGYCJ	naringenin-chalcon
40	47	34.8	388	1 SYGYC3	naringenin-chalcon
41	47	34.8	388	2 S37098	naringenin-chalcon
42	47	34.8	388	2 JQ2249	naringenin-chalcon
43	47	34.8	388	2 JQ2259	naringenin-chalcon
44	47	34.8	388	2 S60472	naringenin-chalcon
45	47	34.8	388	2 JCS516	naringenin-chalcon

ALIGNMENTS

RESULT 1
KABS02
alpha-s2-casein precursor - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 01-Sep-1981 #sequence revision 03-Feb-1994 #text_change 22-Jun-1999
C/Accession: JQ2008; A29087; A91438; S66626; A03107
R/Groenen, M.A.M.; Dijkhof, R.J.M.; Verstege, A.T.M.; van der Poel, J.J.
Gene 123, 187-193, 1993
A/Title: The complete sequence of the gene encoding bovine alpha2-casein.
A/Reference number: JQ2008; MUID:93154583; PMID:8428658
A/Accession: JQ2008
A/Status: translation not shown
A/Molecule type: DNA
A/Residues: 1-222 <GRO>
A/Cross-references: GB:M644; NID:9162928; PID:AAA30479.1; PID:9162929
R/Stewart, A.F.; Bonsing, J.; Beattie, C.W.; Shah, F.; Willis, I.M.; Mackinlay, A.G.
Mol. Biol. Evol. 4, 231-241, 1987
A/Title: Complete nucleotide sequences of bovine alpha-s2- and beta-casein cDNAs: compar
A/Reference number: A93062; MUID:88188989; PMID:2833669
A/Accession: A29087
A/Status: translation not shown
A/Molecule type: mRNA
A/Residues: 1-222 <STB>
A/Cross-references: GB:M644; NID:9162928; PID:AAA30479.1; PID:9162929
R/Birionon, G.; Ribadeau-Dumas, B.; Mercier, J.C.; Pelissier, J.P.; Des, B.C.
FEBS Lett. 76, 274-279, 1977
A/Title: Complete amino acid sequence of bovine alpha-s2-casein.
A/Reference number: A91438; MUID:77185633; PMID:862906
A/Accession: A91438
A/Molecule type: protein
A/Residues: 16-101, 'EE', 104-222 <BRI>
A/Note: four fractions, previously designated s2, s3, s4, and s6, appear to have the sam
these
J. Grosclaude, F.; Joudrier, P.; Mahe, M.F.
J. Dairy Res. 46, 211-213, 1979
A/Title: A genetic and biochemical analysis of a polymorphism of bovine alpha-s2-casein
A/Reference number: A92771; MUID:79239637; PMID:469044
A/Contents: annotation; D allele
A/Note: the sequence of the D allele has a deletion of nine residues, which may be 49-58
R/Zucht, H.D.; Raida, M.; Adermann, K.; Maegert, H.J.; Forstmann, W.G.
FEBS Lett. 372, 185-188, 1995
A/Title: Casein alpha-2: a casein-alpha(s2) derived peptide exhibits antibacterial activity
A/Reference number: S66626; MUID:9600204; PMID:755666
A/Accession: S66626
A/Molecule type: protein
A/Residues: 165-203 <ZUC>
A/Comment: The sequence of the A allele is shown.
C/Genetics:
A/Gene: alpha2s2a
A/Map position: 6
A/Introns: 17/3; 26/3; 33/3; 47/3; 56/3; 65/3; 74/3; 82/3; 97/3; 138/3; 147/3; 156/3; 16

C/Superfamily: alpha-s2-casein
 C/Keywords: mammary gland; milk; phosphoprotein
 F/1-15/Domain: signal sequence #status predicted <SIG>
 F/16-223/Product: alpha-s2-casein #status experimental <MAT>
 F/23,24,25,31,71,72,73,76,144,146,158/Binding site: phosphate (Ser) (covalent) #status F

Query Match 100.0%; Score 135; DB 1; Length 222;
 Best Local Similarity 100.0%; Pred. No. 1,1e-12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKTVIPYRY 24
 |||||
 Db 198 VYOHOKAMKPMIOPKTVIPYRY 221

RESULT 2

KASHS2
 alpha-s2-casein precursor - sheep
 C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C/Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 22-Jun-1999
 C/Accession: A25070; S17856
 R/Bouinard, M.; Petrisant, G.
 Biochimie 67, 1043-1051, 1985
 A/Title: Complete sequence of ovine alpha-s2-casein messenger RNA.
 A/Reference number: A25070; MUID:86104467; PMID:3002499
 A/Accession: A25070
 A/Molecule type: mRNA
 A/Residues: 1-223 <BO1>

A/Cross-references: GB:X03238; NID:g1238; PIDN:CMA46993.1; PID:G732894
 A/Note: 64 Asn was also found
 R/Bouinard, M.; Hue, D.; Bouinard, C.; Mercier, J.C.; Gaye, P.
 Eur. J. Biochem. 201, 633-641, 1991
 A/Title: Multiple mRNA species code for two non-allelic forms of ovine alpha-s2-casein.
 A/Reference number: S17856; MUID:92036619; PMID:1935959

A/Accession: S17856
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-93, 'T', 95-223 <BO2>
 C/Superfamily: alpha-s2-casein
 C/Keywords: mammary gland; milk; phosphoprotein
 F/1-15/Domain: signal sequence #status predicted <SIG>
 F/16-223/Product: alpha-s2-casein #status predicted <KA2>
 F/23,24,25,32,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status F/53,88,146,154,170,198/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 82.2%; Score 111; DB 1; Length 223;
 Best Local Similarity 83.3%; Pred. No. 4,1e-09;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKTVIPYRY 24
 |||||
 Db 199 VDOHOKAMKPMIOPKTVIPYRY 222

RESULT 3

UN0547
 alpha-s2-casein precursor - goat
 C/Species: Capra aegagrus hircus (domestic goat)
 C/Date: 31-Dec-1993 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C/Accession: S33881; S33880; UN0547; S20620
 R/Bouinard, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Protein Seq. Data Anal. 5, 213-218, 1993
 A/Title: Characterization of goat allelic alpha-s2-caseins A and B: further evidence of

A/Accession: S33881
 A/Status: preliminary
 A/Molecule type: protein
 A/Residues: 1-223 <BOU>
 A/Note: allele B
 A/Accession: S33880
 A/Molecule type: protein
 A/Residues: 15-78, 'E', 80-223 <BO2>
 A/Note: sequence deduced from compositional analysis of peptides

R/Bouinard, C.
 Gene 125, 235-236, 1993
 A/Title: Sequence of the goat alpha-s2-casein-encoding cDNA.
 A/Reference number: UN0547; MUID:93216130; PMID:8462880

A/Accession: UN0547
 A/Molecule type: mRNA
 A/Residues: 1-78, 'E', 80-223 <BO3>
 A/Cross-references: EMBL:X65160; NID:g955; PIDN:CMA46278.1; PID:g956
 A/Note: allele A
 C/Superfamily: alpha-s2-casein

C/Keywords: mammary gland; milk; phosphoprotein
 F/1-15/Domain: signal sequence #status predicted <SIG>
 F/16-223/Product: alpha-s2-casein #status experimental <MAT>
 F/23,24,25,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status F/53,146/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 82.2%; Score 111; DB 2; Length 223;
 Best Local Similarity 83.3%; Pred. No. 4,1e-09;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKTVIPYRY 24
 |||||
 Db 199 VDOHOKAMKPMIOPKTVIPYRY 222

RESULT 4

I46995
 alpha-s2-casein C - goat
 C/Species: Capra aegagrus hircus (domestic goat)
 C/Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 13-Aug-1999
 C/Accession: I46995
 R/Bouinard, C.; Brignon, G.; Mahe, M.F.; Printz, C.
 Anim. Genet. 25, 173-177, 1994
 A/Title: Biochemical and genetic analysis of variant C of caprine alpha-s2-casein (Capr-
 A/Reference number: I46995; MUID:95030556; PMID:7943551
 A/Accession: I46995
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-223 <BOU>
 A/Cross-references: GB:S74171; NID:g707033; PIDN:AAB32166.1; PID:g707034
 C/Superfamily: alpha-s2-casein

Query Match 82.2%; Score 111; DB 2; Length 223;
 Best Local Similarity 83.3%; Pred. No. 4,1e-09;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMIOPKTVIPYRY 24
 |||||
 Db 199 VDOHOKAMKPMIOPKTVIPYRY 222

RESULT 5

A48383
 alpha-s2-casein - pig
 C/Species: Sus scrofa domestica (domestic pig)
 C/Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
 C/Accession: A48383
 R/Alexander, L.J.; Das Gupta, N.A.; Beattie, C.W.
 Anim. Genet. 23, 365-367, 1992
 A/Title: The sequence of porcine alpha-s2-casein cDNA.
 A/Reference number: A48383; MUID:92367960; PMID:1503276
 A/Accession: A48383
 A/Status: preliminary
 A/Molecule type: nucleic acid
 A/Residues: 1-235 <ALB>
 A/Experimental source: mammary gland
 A/Note: sequence inconsistent with the nucleotide translation
 A/Note: sequence extracted from NCBI backbone (NCBIN:110884, NCBIPI:110885)
 C/Superfamily: alpha-s2-casein

Query Match 52.2%; Score 70.5; DB 2; Length 235;
 Best Local Similarity 55.6%; Pred. No. 0.005;
 Matches 15; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

A:Gene: chsl
A:Map position: V
A:Note: chsl is expressed in various floral tissues and UV illuminated seedlings
C:Superfamily: chalcone synthase
C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKEMGQPKSKI 124

RESULT 11

SYRUCN
naringenin-chalcone synthase (EC 2.3.1.74) R - garden petunia

N:Alternate names: chalcone synthase
C:Species: Petunia x hybrida (garden petunia)

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
C/Accession: A23643

R/Koe: R.E.; Spelt, C.E.; Reif, H.J.; van den Elzen, P.J.M.; Veltkamp, E.; Mol, J.N.M.
Nucleic Acids Res. 14, 5229-5239, 1986

A>Title: Floral tissue of Petunia hybrida (V30) expresses only one member of the chalcone synthase gene family

A/Reference number: A23643; MUID:86286540; PMID:3016642

A/Accession: A23643

A/Molecule type: mRNA

A/Residues: 1-389 <KOE>

A/Cross-references: GB:X04080; NID:G20541; PIDN:CAA2718.1; PID:G20542

A:Experimental source: strain Violet 30, flowers

C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoids

A:Gene: chsl

A:Note: expressed in floral tissue

C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKEMGQPKSKI 124

RESULT 12

SYRUCN
naringenin-chalcone synthase (EC 2.3.1.74) A - garden petunia

N:Alternate names: chalcone synthase
C:Species: Petunia x hybrida (garden petunia)

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
C/Accession: J03038

R/Koe: R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.
Gene 81, 245-257, 1989

A>Title: Cloning and molecular characterization of the chalcone synthase multigene family

A/Reference number: J03038; MUID:50034197; PMID:2806915

A/Accession: J03038

A/Molecule type: DNA

A/Residues: 1-389 <KOE>

A/Cross-references: GB:X14591; NID:G20524; PIDN:CAA32731.1; PID:G20525

A:Experimental source: strain Violet 30, leaf

C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoids

A:Gene: chsl

A:Note: expressed in various floral tissues and UV illuminated seedlings

C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 38.5%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;

Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKEMGQPKSKI 124

Query Match 38.5%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKEMGQPKSKI 124

RESULT 13

SYRUCN
naringenin-chalcone synthase (EC 2.3.1.74) 2 - potato

N:Alternate names: chalcone synthase; CHS
C:Species: Solanum tuberosum (potato)

C>Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 05-May-2000
C/Accession: J05136; PC4239

R/Koe: J.H.; Kim, H.S.; Choi, K.H.; Jung, Y.H.; Jung, H.; Byun, S.M.
Biochem. Biotechnol. 60, 1907-1910, 1996

A>Title: Cloning and characterization of one member of the chalcone synthase gene family

A/Reference number: J05136; MUID:97141614; PMID:987872

A/Accession: J05136

A/Molecule type: mRNA

A/Residues: 1-389 <JEO1>

A/Cross-references: GB:U47738; NID:G1470059; PIDN:AB05239.1; PID:G1470060

A:Status: preliminary

A/Molecule type: protein

A/Residues: 158-165;367-373 <JEO2>

C:Comment: This enzyme is important in the biosynthesis of all classes of flavonoids in C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A

Query Match 38.5%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKEMGQPKSKI 124

RESULT 14

SYRUCN
naringenin-chalcone synthase (EC 2.3.1.74) - common morning-glory

N:Alternate names: chalcone synthase
C:Species: Ipomoea purpurea (common morning-glory)

C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000
C/Accession: T07799

R/Koe: Fukada-Tanaka, S.; Hoshino, A.; Hisatomi, Y.; Habu, Y.; Hasebe, M.; Iida, S.
Plant Cell Physiol. 38, 754-758, 1997

A>Title: Identification of new chalcone synthase genes for flower pigmentation in the J

A/Reference number: Z16140; MUID:97393496; PMID:9249990

A/Accession: T07799

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-389 <FUK>

A/Cross-references: EMBL:AB001827; NID:G2329836; PIDN:BAA21789.1; PID:G2329837

C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoids

A:Gene: CHD-E

C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 38.5%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 OKAMKPMIOPKTV 18
DB 111 OKAIKEMGQPKSKI 124

RESULT 15

SYRUCN
naringenin-chalcone synthase (EC 2.3.1.74) - garden snapdragon

N/Alternate names: chalcone synthase
C/Species: Antirrhinum majus (garden snapdragon)
C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000
C/Accession: S07312; A33217
R/Sommer, H.; Saedler, H.
Mol. Gen. Genet. 202, 429-434, 1986
A/Title: Structure of the chalcone synthase gene of Antirrhinum majus.
A/Reference number: S07312
A/Accession: S07312
A/Molecule type: DNA
A/Residues: 1-390 <SOM>
A/Cross-references: EMBL:X03710; NID:G16015; PTDN:CAA27338.1; PID:G16016
A/Accession: A33217
A/Status: translation not shown
A/Molecule type: mRNA
A/Residues: 1-390 <SOM2>
C/Genetics:
A/Gene: chs
A/Introns: 60/1, 162/3
C/Superfamily: chalcone synthase
C/Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 38.5%; Score 52; DB 1; Length 390;
Best Local Similarity 64.3%; Pred. No. 5;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIQPKTKV 18
|||:| | |||:
Db 111 OKAIKEMGQPKSKI 124

Search completed: July 30, 2003, 16:31:15
Job time : 15.2857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 47.2857 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-3

Perfect score: 135

Sequence: 1 VYQHOKAMKPMIQPKTKVPIYVRY 24

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSI/gcgdata/geneseq/emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	135	100.0	24	15	AA60481
2	135	100.0	24	21	AA788339
3	135	100.0	25	13	AA20055
4	135	100.0	25	21	AA788340
5	135	100.0	31	18	AAW32219
6	135	100.0	31	23	AA17466
7	135	100.0	222	18	AAW32220
8	135	100.0	222	23	AA17468
9	124.5	92.2	26	20	AAW93885

10	119	88.1	21	21	AAV88338	Internal peptide f
11	111	82.2	223	23	AAE17469	Alpha-S2 casein pr
12	111	82.2	223	23	AAE17470	Goat alpha-S2 case
13	111	82.2	223	23	AAE17471	Capra hircus alpha
14	111	82.2	223	23	AAE17472	Sheep alpha-S2 cas
15	111	82.2	223	23	AAE17473	Alpha-S2 casein pr
16	101	74.8	19	18	AAW32218	Alpha-S2 casein pr
17	101	74.8	19	23	AAE17465	Bovine alpha-S2 ca
18	92	68.1	16	21	AAV88337	Internal peptide f
19	70.5	52.2	234	23	AAE17474	Pig alpha-S2 casei
20	67	49.6	178	23	AAE17472	Rabbit alpha-S2 ca
21	65	48.1	13	23	AAE17464	Bovine alpha-S2 ca
22	60	44.4	12	23	AAE17463	Bovine alpha-S2 ca
23	53	39.3	9	17	AAW04282	Milk derived anti-
24	53	39.3	11	18	AAW32217	Alpha-S2 casein pr
25	53	39.3	11	23	AAE17462	Bovine alpha-S2 ca
26	52	38.5	360	22	AAE60169	Petunia chalcone s
27	52	38.5	389	17	AAE60125	Chalcone synthase.
28	52	38.5	389	22	AAE74019	Cyclamen chalcone
29	52	38.5	389	24	ABG73901	Petunia chalcone s
30	51	37.8	394	24	ABE60374	Humulus lupulus ch
31	51	37.8	399	24	ABE60375	Humulus lupulus ch
32	50	37.0	291	22	AAE60992	Mouse GN protein s
33	49	36.3	8	17	AAW04278	Milk derived anti-
34	49	36.3	133	17	AAW04828	Vascular endotheli
35	49	36.3	133	20	AAW80492	Murine vascular en
36	49	36.3	133	23	AAU83406	Vascular endotheli
37	49	36.3	160	20	AAW86208	Human VEGF-B trunc
38	49	36.3	167	20	AAW86234	Human VEGF-B full
39	49	36.3	188	17	AAW04826	Heart vascular end
40	49	36.3	188	17	AAW00864	Murine VRF167. Mu
41	49	36.3	188	20	AAW86201	Human vascular en
42	49	36.3	188	20	AAW80490	Murine vascular en
43	49	36.3	188	23	AAU83404	Murine vascular en
44	49	36.3	195	17	AAW04827	Heart vascular end
45	49	36.3	195	20	AAW80491	Murine vascular en

ALIGNMENTS

RESULT 1
AA60481
ID AA60481 standard; peptide; 24 AA.
AC AA60481;
DT 19-APR-1995 (first entry)
XX
DB Peptide used in tranquilliser.
XX
KW Peptide; tranquilliser; food; foodstuff; mental stabilisation.
OS Synthetic.
XX
PN JP06211689-A.
XX
PD 02-AUG-1994.
XX
PF 19-JAN-1993; 93JP-0024811.
XX
PR 19-JAN-1993; 93JP-0024811.
XX
PA (KANE) KANEBO LTD.
XX
DR WPI; 1994-283276/35.
XX
PT Synthetic peptide(s) used in tranquilliser - also used in the
production of a food for mental stabilisation.
XX
PS Claim 3; Page 1; 5pp; Japanese.
XX
CC The peptide may be used as a component of a tranquilliser which may

CC be administered orally and is low in toxicity. The peptide may also
 CC be used as a component of a foodstuff which may be used for mental
 CC stabilisation. See also AAR60480.

SO Sequence 24 AA;

Query Match 100.0%; Score 135; DB 15; Length 24;
 Best Local Similarity 100.0%; Pred. No. 6.2e-13;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWIOPTKTVIPYRY 24
 1 VYOHQKMKPWIOPTKTVIPYRY 24
 Db 1 VYOHQKMKPWIOPTKTVIPYRY 24

RESULT 2

AAV88339

14-JUL-2000 (first entry)

Internal peptide fragment of bovine alpha-S2 caesin #3.

Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;

antibacterial peptide; lactoferrin; antiviral; antitumour activity.

Bos sp.

WO200015655-A1.

23-MAR-2000.

15-SEP-1999; 99WO-EP07002.

15-SEP-1998; 98EP-0203107.

08-JUN-1999; 99EP-0201815.

(NIZO-) NIZO FOOD RES.

Visser S, Reclot I;

WPI; 2000-271377/23.

Novel process for producing peptides with e.g. antimicrobial activity
 from biological fluids such as milk, whey or blood comprises contacting
 fluid with chromatographic medium to adsorb peptide domain of interest

Claim 14; Page 22; 41pp; English.

This sequence represents an internal fragment of bovine alpha-S2 caesin
 protein. The peptide is an example of a peptide with antibacterial
 activity that can be produced by the process of the invention. The
 invention relates to a process for producing peptides from biological
 fluids. The process comprises chromatography of the biological fluid, in
 situ hydrolysis of selectively bound peptides, washing to remove unbound
 peptide, and elution of the peptides of interest. The process is used for
 producing peptides from biological fluids, such as milk, whey or blood.
 For example, the process can be used to produce antibacterial peptides
 derived from lactoferrin, using cheese whey as a starting material. The
 peptides obtained have preferably antimicrobial and/or antiviral and/or
 antitumour activity. The process of the invention is relatively simple
 and generally economically and technically more attractive than those
 methods previously used. The method provides high yield peptides with a
 selected activity of interest without the need for intermediate
 purification of the precursor protein.

SO Sequence 24 AA;

Query Match 100.0%; Score 135; DB 21; Length 24;
 Best Local Similarity 100.0%; Pred. No. 6.2e-13;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWIOPTKTVIPYRY 24
 1 VYOHQKMKPWIOPTKTVIPYRY 24
 Db 1 VYOHQKMKPWIOPTKTVIPYRY 24

RESULT 3
 AAR20055
 ID AAR20055 standard; peptide; 25 AA.

AC AAR20055;

25-MAR-2003 (updated)
 DT 26-MAR-1992 (first entry)

Casein peptide.

Casein; platelet; aggregation; thrombosis; collagen.

Synthetic.

JF03255095-A.

13-MAR-1991.

02-MAR-1990; 90JP-0052554.

02-MAR-1990; 90JP-0052554.

(KANE) KANEBO LTD.

WPI; 1992-002669/01.

Casein peptide(s) for treating thrombosis - as inhibitors of
 platelet aggregation caused by adenosine-5'-di-phosphate and
 collagen and as biochemical reagents

Claim 1; Page 1; 8pp; Japanese.

Leu25 may be omitted. The peptide and its salts have inhibiting
 activity against platelet aggregation caused by adenosine-5'-
 di-phosphate and by collagen. They are useful for prevention and
 treatment of thrombosis and are also useful as a biochemical reagent.
 (Updated on 25-MAR-2003 to correct PA field.)

SO Sequence 25 AA;

Query Match 100.0%; Score 135; DB 13; Length 25;
 Best Local Similarity 100.0%; Pred. No. 6.5e-13;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYOHQKMKPWIOPTKTVIPYRY 24
 1 VYOHQKMKPWIOPTKTVIPYRY 24
 Db 1 VYOHQKMKPWIOPTKTVIPYRY 24

RESULT 4
 AAY88340
 ID AAY88340 standard; peptide; 25 AA.

AC AAY88340;

14-JUL-2000 (first entry)

C-terminal peptide fragment of bovine alpha-S2 caesin.

Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;

antibacterial peptide; lactoferrin; antiviral; antitumour activity.

Bos sp.

WO200015655-A1.

XX 23-MAR-2000.
 PD 15-SEP-1999; 99WO-EP07002.
 XX 15-SEP-1998; 98EP-0203107.
 PF 08-JUN-1999; 99EP-0201815.
 XX (NIZO-) NIZO FOOD RES.
 XX Visser S, Recto I;
 PI WPI; 2000-271377/23.
 DR Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest
 PT -
 PS Claim 14; Page 22; 41pp; English.
 XX This sequence represents a C-terminal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.
 CC
 SQ Sequence 25 AA;
 OY Query Match 100.0%; Score 135; DB 21; Length 25;
 Best Local Similarity 100.0%; Pred. No. 6.5e-13;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 VYOHOKAMKPMWIOPTKTVIPYRY 24
 1 VYOHOKAMKPMWIOPTKTVIPYRY 24
 RESULT 5
 AAM32219
 ID AAM32219 standard; peptide; 31 AA.
 AC AAM32219;
 XX
 DT 03-FEB-1998 (first entry)
 XX
 DE Alpha-S2 casein precursor C-terminal peptide fragment #5.
 XX
 KW Alpha-S2 casein precursor; growth promoting; mitogenic assay;
 KW platelet-derived growth factor; insulin-like derived growth factor;
 XX
 OS Synthetic.
 OS Bos taurus.
 XX
 PN WO9716460-A1.
 PD 09-MAY-1997.
 XX
 PF 31-OCT-1996; 96WO-GB02658.
 XX
 PR 31-OCT-1995; 95GB-0022302.
 XX

PA (UYLI-) UNIV LIVERPOOL.
 XX
 PI Liu Q, Smith JA, Wilkinson MC;
 XX WPI; 1997-272048/24.
 DR Manufacture of medicament or foodstuff for promoting growth - using
 PT peptide(s) with a sequence identical to the C-terminal end of an
 PT alpha-S2 casein precursor
 XX
 PS Claim 12; Page 21; 33pp; English.
 XX
 CC The present sequence, which is substantially identical to the C-terminal
 CC end of an alpha-S2 casein precursor, was found after storage in PBS to
 CC exhibit growth promoting activity for rat mammary fibroblast cell line
 CC (Rama 27), which is not significantly stimulated by platelet-derived
 CC growth factor or insulin-like growth factor. The activity of the peptide
 CC increased when maintained at alkaline pH. By way of contrast, alpha-S2
 CC casein was inactive in a mitogenic assay. This peptide may be used in the
 CC manufacture of a medicament or foodstuff for promoting growth in
 CC humans or animals.
 CC
 SQ Sequence 31 AA;
 OY Query Match 100.0%; Score 135; DB 18; Length 31;
 Best Local Similarity 100.0%; Pred. No. 8.3e-13;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 7 VYOHOKAMKPMWIOPTKTVIPYRY 30
 1 VYOHOKAMKPMWIOPTKTVIPYRY 24
 7 VYOHOKAMKPMWIOPTKTVIPYRY 30
 RESULT 6
 AAE17466
 ID AAE17466 standard; peptide; 31 AA.
 AC AAE17466;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor protein C-terminal fragment #7.
 XX
 KW Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KW chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Bos sp.
 XX
 PN WO200202133-A2.
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX WPI; 2002-154690/20.
 DR Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 9; Page 22; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and

CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is a peptide
 CC derived from the C-terminal of bovine alpha-S2 casein precursor protein.
 XX
 SQ Sequence 31 AA;

Query Match 100.0%; Score 135; DB 23; Length 31;
 Best Local Similarity 100.0%; Pred. No. 7.3e-12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKTKVPIPVRY 24
 DB 7 VYQHOKAMKPMIOPKTKVPIPVRY 30

RESULT 7

AAW32220 standard; protein; 222 AA.

AC AAW32220;
 DT 03-FEB-1998 (first entry)
 XX

DE Bovine alpha-S2 casein precursor.

KW Bovine alpha-S2 casein precursor; growth promoting; mitogenic assay;
 KM platelet-derived growth factor; insulin-like derived growth factor;
 XX

OS Bos taurus.

PN WO9716460-A1.

PD 09-MAY-1997.

PF 31-OCT-1996; 96WO-GB02658.

PR 31-OCT-1995; 95GB-0022302.

PA (UYLI-) UNIV LIVERPOOL.

PI Liu Q, Smith JA, Wilkinson MC;

DR WPI; 1997-272048/24.

PT Manufacture of medicament or foodstuff for promoting growth - using
 PT peptide(s) with a sequence identical to the C-terminal end of an
 PT alpha-S2 casein precursor

PS Disclosure; Page 3; 33pp; English.

XX The present sequence represents bovine alpha-S2 casein precursor.
 CC Peptides having an amino acid sequence which is substantially identical
 CC to the C-terminal end of an alpha-S2 casein precursor, are used for
 CC manufacture of a medicament or foodstuff for promoting growth in humans
 CC or animals.
 CC

SQ Sequence 222 AA;

Query Match 100.0%; Score 135; DB 18; Length 222;
 Best Local Similarity 100.0%; Pred. No. 7.3e-12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKTKVPIPVRY 24
 DB 198 VYQHOKAMKPMIOPKTKVPIPVRY 221

RESULT 8

AAE17468
 ID AAE17468 standard; Protein; 222 AA.

AC AAE17468;

DT 22-APR-2002 (first entry)

DE Bovine alpha-S2 casein precursor protein.

KW Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX

OS Bos sp.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX

PS Claim 8; Page 6; 27pp; English.

XX The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptide stimulates the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is bovine
 CC alpha-S2 casein precursor protein.
 CC

SQ Sequence 222 AA;

Query Match 100.0%; Score 135; DB 23; Length 222;
 Best Local Similarity 100.0%; Pred. No. 7.3e-12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKTKVPIPVRY 24
 DB 198 VYQHOKAMKPMIOPKTKVPIPVRY 221

RESULT 9

AAW93885
 ID AAW93885 standard; peptide; 26 AA.

AC AAW93885;

DT 25-JUN-1999 (first entry)

DE Bifidobacterium bifidus stimulating peptide 21.

XX Bifidogenic peptide; protease; treatment; microbe-related disease;
 KM Bacteria; fungi; yeast; protozoa; virus; mycoplasma; filaria; plasmodia;
 KM infection; inflammation; microbial induced tumour; degenerative disorder;
 KM diarrhoea; colic; oral microflora; intestinal microflora; caries;
 KM

KM vaginal microflora.
 XX Bifidobacterium bifidus.
 OS
 XX MO9914231-A2.
 PN
 XX 25-MAR-1999.
 PD
 XX 16-SEP-1998; 98WO-EP05899.
 PF
 XX 11-FEB-1998; 98DE-1005385.
 PR
 XX 16-SEP-1997; 97DE-1040604.
 PA
 XX (FORS/) FORSSMANN W.
 PI
 XX Forssmann W, Liepke C, Zucht H;
 DR WPI; 1999-244022/20.
 XX
 XX Milk-derived peptides that stimulate Bifidobacterium bifidus
 PT
 XX
 PS Claim 2; Page 3; 25pp; German.
 XX
 CC This invention describes milk-derived bifidogenic peptides and their
 CC active derivatives or fragments, and combinations of them produced by
 CC chemical coupling. Such are produced from bovine or human milk by
 CC treatment for 2 hr with processes, then centrifuging to remove fat and
 CC acidifying to pH 2 to precipitate proteins. The solution phase is then
 CC subjected to reverse-phase high-performance liquid chromatography (HPLC)
 CC and cation-exchange HPLC, the fractions adjusted to salt content below
 CC 25 mM (by dialysis or reverse-phase HPLC) and tested for activity by
 CC growing Bifidobacterium bifidus and Escherichia coli in presence of the
 CC fractions. Those fractions for which (Bw-B0) -(Bw-E0) is at least 0.15
 CC are selected where Bw = germ count after 16 hr culture of B. bifidus in
 CC 50% Bliker broth containing peptide at 0.2 mg/ml, B0 = germ count under
 CC similar conditions in a peptide-free control, Bw = germ count after 16
 CC hr culture of B. coli in 3 g/l tryptic broth containing peptide at
 CC 0.2 mg/ml, B0 = germ count under similar conditions in a peptide-free
 CC control. The peptides AAW93865-W93888 are used to treat microbe-related
 CC diseases caused by bacteria, fungi, yeast, protozoa, viruses,
 CC mycoplasma, filaria and plasmodia, e.g. infections, inflammation,
 CC microbially induced tumours or degenerative disorders, diarrhoea, colic,
 CC abnormalities in oral, intestinal or vaginal microflora, or caries.
 CC
 XX
 SQ Sequence 26 AA;
 Query Match 92.2%; Score 124.5; DB 20; Length 26;
 Best Local Similarity 96.0%; Pred. No. 2,4e-11;
 Matches 24; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 VYQHOKAM-KPWIOPTKTVIPYRY 24
 DB 1 VYQHOKAMKPWIOPTKTVIPYRY 25
 RESULT 10
 AAY88338
 ID AAY88338 standard; peptide; 21 AA.
 XX
 AC AAY88338;
 XX
 DT 14-JUL-2000 (first entry)
 XX
 DE Internal peptide fragment of bovine alpha-S2 casein #2.
 XX
 KM Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
 KM antibacterial peptide; lactoferrin; antiviral; antitumour activity.
 XX
 OS Bos sp.
 XX
 PN WO200015655-A1.
 XX
 PD 23-MAR-2000.
 XX

XX
 PF 15-SEP-1999; 99WO-EP07002.
 XX
 PR 15-SEP-1998; 98EP-0203107.
 PR 08-JUN-1999; 99EP-0201815.
 XX
 XX (NIZO-) NIZO FOOD RES.
 PA
 XX
 PI Visser S, Recto I;
 XX
 DR WPI; 2000-271377/23.
 XX
 XX Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest
 PT
 XX
 XX Claim 14; Page 22; 41pp; English.
 PS
 XX
 CC This sequence represents an internal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.
 CC
 XX
 SQ Sequence 21 AA;
 Query Match 88.1%; Score 119; DB 21; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.2e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPWIOPTKTVIPY 21
 DB 1 VYQHOKAMKPWIOPTKTVIPY 21
 RESULT 11
 AAE17469
 ID AAE17469 standard; Protein; 223 AA.
 XX
 AC AAE17469;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Alpha-S2 casein precursor (alpha-S2-CN) protein #1.
 XX
 KM Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
 KM fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
 KM toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Unidentified.
 XX
 PN WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX

PI Smith JA;
 XX WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Disclosure; Page 8; 27pp; English.
 CC
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is alpha-S2
 CC casein precursor (alpha-S2-CN) protein.
 CC
 SQ Sequence 223 AA;
 Query Match 82.2%; Score 111; DB 23; Length 223;
 Best Local Similarity 83.3%; Pred. No. 2.5e-08;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VYOHQKMKPWIOPTKTVIPYRY 24
 DB 199 VDOHQKMKPWIOPTKTVIPYRY 222
 |||||
 RESULT 12
 ID AAE17470 standard; Protein; 223 AA.
 XX
 AC AAE17470;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Goat alpha-S2 casein E precursor protein.
 XX
 XX Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 OS WO200202133-A2.
 XX
 PN 10-JAN-2002.
 PD
 XX 13-JUN-2001; 2001WO-GB02601.
 PF
 XX 30-JUN-2000; 2000GB-0016189.
 PR
 XX (PEPS-) PEPSYN LTD.
 PA
 XX Smith JA;
 PT
 XX WPI; 2002-154690/20.
 DR
 XX Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 CC
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein C precursor protein.
 CC
 SQ Sequence 223 AA;
 Query Match 82.2%; Score 111; DB 23; Length 223;

CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein E precursor protein.
 CC
 SQ Sequence 223 AA;
 Query Match 82.2%; Score 111; DB 23; Length 223;
 Best Local Similarity 83.3%; Pred. No. 2.5e-08;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 VYOHQKMKPWIOPTKTVIPYRY 24
 DB 199 VDOHQKMKPWIOPTKTVIPYRY 222
 |||||
 RESULT 13
 ID AAE17471 standard; Protein; 223 AA.
 XX
 AC AAE17471;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Capra hircus alpha-S2 casein C precursor protein.
 XX
 XX Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 OS WO200202133-A2.
 XX
 PN 10-JAN-2002.
 PD
 XX 13-JUN-2001; 2001WO-GB02601.
 PF
 XX 30-JUN-2000; 2000GB-0016189.
 PR
 XX (PEPS-) PEPSYN LTD.
 PA
 XX Smith JA;
 PT
 XX WPI; 2002-154690/20.
 DR
 XX Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 CC
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein C precursor protein.
 CC
 SQ Sequence 223 AA;
 Query Match 82.2%; Score 111; DB 23; Length 223;

Best Local Similarity 83.3%; Pred. No. 2.5e-08;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VVOHQKMKPWTQPKTKVPIPVRY 24
DB 199 VDOHQKMKPWTQPKTKVPIPVRY 222

RESULT 14
AAE17473
ID AAE17473 standard; Protein; 223 AA.

AC AAE17473;

DT 22-APR-2002 (first entry)

DE Sheep alpha-S2 casein precursor protein.

KW Sheep; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
KW collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
XX chewing gum; cosmetic; wrinkling; periodontal disease.

OS Ovis sp.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
PT alpha-S2 casein precursor in the manufacture of a medicament for
PT alleviating or preventing periodontal disease and an effect of aging in
PT skin -

PS Claim 8; Page 9; 27pp; English.

XX The invention relates to a composition comprising a peptide or its
CC derivative. The peptide contains an amino acid sequence from alpha-S2
CC casein precursor. The peptides stimulate the growth of fibroblasts, and
CC thus the synthesis and secretion of collagen. The peptides also stimulate
CC the growth of keratinocytes, which aid in the formation and regeneration
CC of skin surface. The peptide is useful in the manufacture of a medicament
CC in the form of a toothpaste or a chewing gum, for alleviating or
CC preventing periodontal disease and a medicament in the form of a cosmetic
CC composition for alleviating or preventing an effect of aging,
CC particularly wrinkling of the skin. The present sequence is sheep
CC alpha-S2 casein precursor protein.

XX Sequence 223 AA;

Query Match 82.2%; Score 111; DB 23; Length 223;
Best Local Similarity 83.3%; Pred. No. 2.5e-08;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VVOHQKMKPWTQPKTKVPIPVRY 24
DB 199 VDOHQKMKPWTQPKTKVPIPVRY 222

RESULT 15
AAE17475
ID AAE17475 standard; Protein; 223 AA.
XX
AC AAE17475;

XX 22-APR-2002 (first entry)

DT Alpha-S2 casein precursor (alpha-S2-CN) protein #2.

DE Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
KW fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
KW toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.

OS Unidentified.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

PT Use of peptide or its derivative containing an amino acid sequence in
PT alpha-S2 casein precursor in the manufacture of a medicament for
PT alleviating or preventing periodontal disease and an effect of aging in
PT skin -

PS Disclosure; Page 9; 27pp; English.

XX The invention relates to a composition comprising a peptide or its
CC derivative. The peptide contains an amino acid sequence from alpha-S2
CC casein precursor. The peptides stimulate the growth of fibroblasts, and
CC thus the synthesis and secretion of collagen. The peptides also stimulate
CC the growth of keratinocytes, which aid in the formation and regeneration
CC of skin surface. The peptide is useful in the manufacture of a medicament
CC in the form of a toothpaste or a chewing gum, for alleviating or
CC preventing periodontal disease and a medicament in the form of a cosmetic
CC composition for alleviating or preventing an effect of aging,
CC particularly wrinkling of the skin. The present sequence is alpha-S2
CC casein precursor (alpha-S2-CN) protein.

XX Sequence 223 AA;

Query Match 82.2%; Score 111; DB 23; Length 223;
Best Local Similarity 83.3%; Pred. No. 2.5e-08;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VVOHQKMKPWTQPKTKVPIPVRY 24
DB 199 VDOHQKMKPWTQPKTKVPIPVRY 222

Search completed: July 30, 2003, 16:23:27
Job time : 48.2857 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 42.4107 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-4

Perfect score: 139
Sequence: 1 VYOHQKMKPWIOPTKVIPIVRYL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_Organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP Vertebrate:*
14: SP Unclassified:*
15: SP_virus:*
16: SP_bacteriap:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123	88.5	210	6	O62825
2	115	82.7	223	6	Q9MYU6
3	106	76.3	223	6	Q9TQ07
4	106	76.3	223	6	Q9GK07
5	53	38.1	935	3	O94623
6	52	37.4	80	10	O81534
7	52	37.4	85	10	O81399
8	52	37.4	340	10	O96568
9	52	37.4	379	10	O04065
10	52	37.4	386	10	O23923
11	52	37.4	388	10	O9FS37
12	52	37.4	389	10	O8LP19
13	52	37.4	389	10	O43040
14	52	37.4	389	10	O93XP8
15	52	37.4	389	10	O9MSB2
16	52	37.4	390	10	O9LRB2

17	52	37.4	390	10	O48564
18	52	37.4	390	10	O22122
19	51	36.7	394	10	O93V86
20	51	36.7	399	10	O94LW8
21	50.5	36.3	471	5	O15918
22	50	36.0	311	5	O12257
23	50	36.0	693	11	Q9WV66
24	50	36.0	1877	3	Q9USR3
25	49.5	35.6	325	16	Q9K947
26	49	35.3	282	5	O44074
27	49	35.3	342	16	Q8YMO6
28	48.5	34.9	286	10	Q8W2F2
29	48	34.5	183	16	Q9CP51
30	48	34.5	208	16	Q8F8G2
31	48	34.5	225	4	Q9BTR9
32	48	34.5	328	16	O8CP60
33	48	34.5	374	10	O81476
34	48	34.5	392	10	Q93YX5
35	48	34.5	392	10	Q9PEK7
36	48	34.5	392	10	O8LPP4
37	48	34.5	392	10	Q9FRW2
38	48	34.5	392	10	Q9SPW2
39	48	34.5	392	10	O8LPP2
40	48	34.5	392	10	Q9S962
41	48	34.5	392	10	O8LPP3
42	48	34.5	392	10	O94G58
43	48	34.5	392	10	O944W8
44	48	34.5	392	10	O944W7
45	48	34.5	395	10	Q9FR70

ALIGNMENTS

RESULT 1	
O62825	PRELIMINARY; PRT; 210 AA.
AC O62825;	
DT 01-AUG-1998 (TREMBLrel. 07, Created)	
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)	
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE AS2-casein (Fragment).	
OS Bubalus bubalis (Domestic water buffalo).	
OC Buiarysta; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
OC Bovidae; Bovinae; Bubalus.	
OX NCBI_TaxID=89462;	
OR	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=Mammary gland;	
RA Das P., Jain S., Garg L.C.;	
RT "Cloning and nucleotide sequence of cDNA encoding as2-casein in B.	
RT bubalis.";	
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.	
DR EMBL; AJ005431; CA06534.2; -	
DR InterPro; IPR001588; Casein.	
DR Pfam; PF00363; caseins; 2.	
FT NON_TER	
SQ SEQUENCE 210 AA; 24700 MW; 05DEF95963F1132C CRC64;	
Query Match	88.5%; Score 123; DB 6; Length 210;
Best local Similarity	88.0%; Pred. No. 6.9e-11;
Matches 22; Conservative 1; Mismatches 2; Indels 0; Gaps 0;	
OY 1 VYOHQKMKPWIOPTKVIPIVRYL 25	
DB 186 VYOHQKMKPWIOPTKVIPIVRYL 210	
RESULT 2	
O9MYU6	PRELIMINARY; PRT; 223 AA.
ID O9MYU6	
AC O9MYU6;	

DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Alpha s2-casein.
 GN CSN152.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OC NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lagomigro R., Pilla F., Mataasino D., Zullo A.;
 RT "A new allele of goat alpha s2-casein gene."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ289716; CAB94236.1; -
 DR InterPro: IPR001588; Casein.
 DR Pfam: PF00363; caseins; 2.
 DR PROSITE: PS00306; CASEIN ALPHA_BETA, 1.
 SQ SEQUENCE 223 AA; 26403 MW; 0E1FE83F24DA85E2 CRC64;

Query Match 82.7%; Score 115; DB 6; Length 223;
 Best Local Similarity 84.0%; Pred. No. 1.2e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VYOHOKAMKPMIOPTKVIPIVRYL 25
 Db 199 VDOHOKAMKPMIOPTKVIPIVRYL 223

RESULT 3

Q9TTO7 PRELIMINARY; PRT; 223 AA.
 AC Q9TTO7;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Alpha s2-casein.
 GN CSN152.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OC NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veltri C.C., Pilla F.F., Lagomigro R.R.;
 RT "A new allele of goat alpha s2-casein."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ249995; CAB59920.1; -
 DR InterPro: IPR001588; Casein.
 DR Pfam: PF00363; caseins; 2.
 DR PROSITE: PS00306; CASEIN ALPHA_BETA, 1.
 SQ SEQUENCE 223 AA; 26433 MW; CE9F4DC8D7688293 CRC64;

Query Match 76.3%; Score 106; DB 6; Length 223;
 Best Local Similarity 80.0%; Pred. No. 3e-08;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VYOHOKAMKPMIOPTKVIPIVRYL 25
 Db 199 VDOHOKAMKPMIOPTKVIPIVRYL 223

RESULT 4

Q9GK07 PRELIMINARY; PRT; 223 AA.
 AC Q9GK07;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Alpha s2-casein.
 GN CSN152.

OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OC NCBI_TaxID=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Veltri C., Pilla F., Lagomigro R.;
 RT "A new allele of alpha s2-casein."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ297310; CAC21704.2; JOINED.
 DR EMBL; AJ297311; CAC21704.2; JOINED.
 DR EMBL; AJ242728; CAC21704.2; JOINED.
 DR EMBL; AJ297312; CAC21704.2; JOINED.
 DR EMBL; AJ297313; CAC21704.2; JOINED.
 DR EMBL; AJ297314; CAC21704.2; JOINED.
 DR EMBL; AJ242527; CAC21704.2; JOINED.
 DR EMBL; AJ297315; CAC21704.2; JOINED.
 DR EMBL; AJ297316; CAC21704.2; JOINED.
 DR EMBL; AJ242526; CAC21704.2; JOINED.
 DR EMBL; AJ242528; CAC21704.2; JOINED.
 DR EMBL; AJ242533; CAC21704.2; JOINED.
 DR InterPro: IPR001588; Casein.
 DR Pfam: PF00363; caseins; 2.
 DR PROSITE: PS00306; CASEIN ALPHA_BETA, 1.
 SQ SEQUENCE 223 AA; 26432 MW; CE9765E8D7688C9D CRC64;

Query Match 76.3%; Score 106; DB 6; Length 223;
 Best Local Similarity 80.0%; Pred. No. 3e-08;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 VYOHOKAMKPMIOPTKVIPIVRYL 25
 Db 199 VDOHOKAMKPMIOPTKVIPIVRYL 223

RESULT 5

O94623 PRELIMINARY; PRT; 935 AA.
 AC O94623;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical 106.5 kDa protein C1347.01C in chromosome II.
 GN SPBC1347.01C OR SPBC215.16C.
 OS Schizosaccharomyces pombe (Pission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 731-935 FROM N.A.
 RC STRAIN=972;
 RA Lyne M., Rajandream M.A., Barrell B.G., Rieger M.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- SIMILARITY: TO C.ELEGANS ZK675.2.
 CC -1- SIMILARITY: BELONGS TO THE UMC (DNA REPAIR) FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BRCT DOMAIN.
 DR EMBL; AL035548; CAB37432.1; -
 DR EMBL; AL035534; CAA22130.1; -
 DR HSSP; P96022; 1IM4.
 DR GeneDB SPombe; SPBC1347.01c; -
 DR InterPro: IPR001357; BRCT.
 DR InterPro: IPR001126; UMC_1like.
 DR Pfam: PF00533; BRCT; 1.
 DR Pfam: PF00817; IMS; 1.
 DR SMART; SM00292; BRCT; 1.

RA Gang Z., Yamazaki M., Sugiyama M., Tanaka Y., Saito K.;
 RT "Cloning and molecular analysis of structural genes involved in
 RT anthocyanin biosynthesis and expressed in a forma-specific manner in
 RT *Perilla frutescens*.";
 RL Plant-Mol. Biol. 35:915-927(1997).
 DR EMBL: AB002582; BAI19548.1; -.
 DR HSSP: P30074; 1D6F.
 DR InterPro: IPR001099; N-C_synthase.
 DR Pfam: PF00195; Chal_stc1_syntC; 1.
 DR Pfam: PF02797; Chal_stc1_syntC; 1.
 DR ProDom: PD000453; N-C_synthase; 1.
 DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 379 AA; 41380 MW; 5E149C5530BD106 CRC64;
 QY
 Db 5 OKAMKPMIOPKTV 18
 111 OKAIKMGQPKSKI 124
 Query Match 37.4%; Score 52; DB 10; Length 379;
 Best Local Similarity 64.3%; Pred. No. 9.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 RESULT 10
 ID 023923 PRELIMINARY; PRT; 386 AA.
 AC 023923;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase (Fragment).
 OS Digitalis lanata (Foxglove).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Lamiales; Antirrhinaceae; Digitalis.
 OC NCBI_TaxID=9450;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RA Thoeninger C., Lindemann P., Luckner M.;
 RT "Expression of phenylalanine ammonia lyase, chalcone synthase and
 RT phytochrome in somatic embryogenesis of *Digitalis lanata*.";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ002526; CAA05512.1; -.
 DR HSSP: P30074; 1D6F.
 DR InterPro: IPR001099; N-C_synthase.
 DR Pfam: PF00195; Chal_stc1_syntC; 1.
 DR Pfam: PF02797; Chal_stc1_syntC; 1.
 DR ProDom: PD000453; N-C_synthase; 1.
 DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 FT NON TER 1 1
 SQ SEQUENCE 386 AA; 42190 MW; 0A8B88FEDA82F8F CRC64;
 QY
 Db 5 OKAMKPMIOPKTV 18
 106 OKAIKMGQPKSKI 119
 Query Match 37.4%; Score 52; DB 10; Length 386;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 RESULT 11
 ID 09FS37 PRELIMINARY; PRT; 388 AA.
 AC 09FS37;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 OS *Torenia hybrida*.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Lamiales; Lamiales incertae sedis; Torenieae;
 OC *Torenia*.
 OC NCBI_TaxID=75807;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Summerwave Blue; TISSUE=Petal;
 RA Suzuki K., Xue H., Tanaka Y., Fukui Y., Fukuchi-Mitani M.,
 RA Murakami Y., Katsumoto Y., Tada S., Kusumi T.;
 RT "flower color modifications of *Torenia hybrida* by cosuppression of
 RT anthocyanin biosynthesis genes.";
 RL Mol. Breed. 6:239-246(2000).
 DR EMBL: AB012923; BAB20074.1; -.
 DR HSSP: P30074; 1D6F.
 DR InterPro: IPR001099; N-C_synthase.
 DR Pfam: PF00195; Chal_stc1_syntC; 1.
 DR Pfam: PF02797; Chal_stc1_syntC; 1.
 DR ProDom: PD000453; N-C_synthase; 1.
 DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 388 AA; 42402 MW; BD3A78B6C44AB0F CRC64;
 QY
 Db 5 OKAMKPMIOPKTV 18
 111 OKAIKMGQPKSKI 124
 Query Match 37.4%; Score 52; DB 10; Length 388;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 RESULT 12
 ID 08LP19 PRELIMINARY; PRT; 389 AA.
 AC 08LP19;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Chalcone synthase.
 OS *Nierembergia* sp. NB17.
 GN CHS.
 ON Nierembergia sp. NB17.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nierembergia.
 OC NCBI_TaxID=184877;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NB17;
 RA Ueyama Y., Katsumoto Y., Fukui Y., Ohkawa H., Kusumi T., Tanaka Y.;
 RT "Flower color modification of *Nierembergia* sp. by engineering
 RT flavonoid biosynthetic pathway.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB078515; BAC10998.1; -.
 DR InterPro: IPR001099; N-C_synthase.
 DR Pfam: PF00195; Chal_stc1_syntC; 1.
 DR Pfam: PF02797; Chal_stc1_syntC; 1.
 DR ProDom: PD000453; N-C_synthase; 1.
 DR PROSITE: PS00441; CHALCONE_SYNTH; 1.
 SQ SEQUENCE 389 AA; 42616 MW; 62DA0C7F407729E7 CRC64;
 QY
 Db 5 OKAMKPMIOPKTV 18
 111 OKAIKMGQPKSKI 124
 Query Match 37.4%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 RESULT 13
 ID 043040 PRELIMINARY; PRT; 389 AA.
 AC 043040;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
 DE Chalcone synthase.
 GN CHS-A.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 NCBI_TaxID=4102;
 RX MEDLINE=96156377; PubMed=8562849;
 RA Shao L., Li Y., Pan A., Cheng Z., Chen M.;
 RT "Molecular cloning, sequencing, and expression in *Escherichia coli* of
 the chalcone synthase gene."
 RL Chin. J. Biotechnol. 11:131-135(1995).
 DR EMBL; F80857; AAB36038.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_stc1_synth; 1.
 DR Pfam; PF02797; Chal_stc1_synth; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCON SYNTH; 1.
 SQ SEQUENCE 389 AA; 42642 MW; A931CF5B255A0A20 CRC64;

Query Match 37.4%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
 DB 111 OKAIKEMGQPKSKI 124

RESULT 14
 ID 093XP8 PRELIMINARY; PRT; 389 AA.
 AC 093XP8;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, last annotation update)
 DE Chalcone synthase.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
 NCBI_TaxID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Flower;
 RA Hu H., Sung H., Su J.;
 RT "Cloning and Expression of a Tobacco Chalcone Synthase Gene."
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF311783; AAK49457.1; -.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_stc1_synth; 1.
 DR Pfam; PF02797; Chal_stc1_synth; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCON SYNTH; 1.
 SQ SEQUENCE 389 AA; 42563 MW; 96C3D426786BD44 CRC64;

Query Match 37.4%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
 DB 111 OKAIKEMGQPKSKI 124

RESULT 15
 Q9MSB2 PRELIMINARY; PRT; 389 AA.
 ID Q9MSB2
 AC Q9MSB2;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
 DE Chalcone synthase (EC 2.3.1.74).
 GN CHS.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 NCBI_TaxID=4102;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Violet; TISSUE=Corolla;
 RA Hsu Y.H., Durdan S.F., To K.Y.;
 RT "Cloning of a pigment-related cDNA encoding chalcone synthase,
 RT chalcone isomerase, dihydroflavonol-4-reductase and cytochrome b5 from
 Petunia hybrida."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF233638; AAF60297.1; -.
 DR HSSP; P30074; 1D6F.
 DR InterPro; IPR001099; N-C_synthase.
 DR Pfam; PF00195; Chal_stc1_synth; 1.
 DR Pfam; PF02797; Chal_stc1_synth; 1.
 DR ProDom; PD000453; N-C_synthase; 1.
 DR PROSITE; PS00441; CHALCON SYNTH; 1.
 KW Acyltransferase; Transferase.
 SQ SEQUENCE 389 AA; 42582 MW; 66135700259D7F0C CRC64;

Query Match 37.4%; Score 52; DB 10; Length 389;
 Best Local Similarity 64.3%; Pred. No. 10;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
 DB 111 OKAIKEMGQPKSKI 124

Search completed: July 30, 2003, 16:29:23
 Job time : 43.4107 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 30, 2003, 16:10:10 ; Search time 8.18452 Seconds

(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-4

Perfect score: 139
Sequence: 1 VYOHQKAMKPMIQPKTVIPYRYL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	139	100.0	222	1 CAS2_BOVIN	P02663 bos taurus
2	115	82.7	223	1 CAS2_CAPII	P33049 capra hircu
3	115	82.7	223	1 CAS2_SHEEP	P04654 ovis aries
4	70.5	50.7	235	1 CAS2_PIG	P39036 sus scrofa
5	67	48.2	182	1 CAS3_RABIT	P50419 oryctolagus
6	55	39.6	193	1 CAS2_CAMDR	O97944 camelus dro
7	52	37.4	389	1 CHS1_LYCES	P23418 lycopersico
8	52	37.4	389	1 CHS2_LYCES	P23419 lycopersico
9	52	37.4	389	1 CHS2_SOLUT	P08894 petunia tub
10	52	37.4	389	1 CHS2_PETTH	O41436 solanum tub
11	52	37.4	389	1 CHS2_PETTH	O41436 solanum tub
12	52	37.4	389	1 CHS2_SOLUT	O41436 solanum tub
13	52	37.4	389	1 CHS2_PETTH	O22047 ipomoea pur
14	52	37.4	389	1 CHS2_PETTH	O22046 pharbitis n
15	52	37.4	389	1 CHS2_PETTH	P22928 petunia hyb
16	52	37.4	390	1 CHS2_PETTH	P06515 antirrhinum
17	52	37.4	391	1 CHS2_PETTH	O04111 petilla fru
18	52	37.4	400	1 CHS2_PETTH	O94941 sorghum bic
19	52	37.4	401	1 CHS1_SORBI	O94942 sorghum bic
20	52	37.4	401	1 CHS2_SORBI	O94943 sorghum bic
21	52	37.4	401	1 CHS3_SORBI	O94944 sorghum bic
22	52	37.4	401	1 CHS4_SORBI	O94945 sorghum bic
23	52	37.4	401	1 CHS5_SORBI	O94946 sorghum bic
24	51	36.7	180	1 CAS2_PABIT	P50418 oryctolagus
25	50	36.0	291	1 PCTL_MOUSE	O94947 mus musculu
26	49.5	35.6	564	1 YSC6_STRGC	P49766 mus musculu
27	49	35.3	392	1 VEBH_MOUSE	P53414 secale cere
28	49	35.3	392	1 CHS1_SECC	P53415 secale cere
29	49	35.3	394	1 CHS2_SECC	P26018 hordeum vul
30	49	35.3	398	1 CHS1_HORVU	P24825 zea mays (m
31	49	35.3	400	1 CHS2_MAIZE	P51071 vitis vinif
32	48	34.5	389	1 THS3_VITVI	P28343 vitis vinif
33	48	34.5	392	1 THS1_VITVI	P28343 vitis vinif

34	48	34.5	392	1 THS2_VITVI	P51070 vitis vinif
35	48	34.5	5171	1 BPEA_HUMAN	O94833 homo sapien
36	47	33.8	285	1 CHS6_MEDSA	P51079 medicago sa
37	47	33.8	311	1 CHS4_TRISU	P51086 trifolium s
38	47	33.8	388	1 CHS1_SOYBN	P24826 glycine max
39	47	33.8	388	1 CHS2_SOYBN	P17957 glycine max
40	47	33.8	388	1 CHS3_SOYBN	P19168 glycine max
41	47	33.8	388	1 CHS5_SOYBN	P48406 glycine max
42	47	33.8	388	1 CHS6_SOYBN	P30080 glycine max
43	47	33.8	388	1 CHS7_SOYBN	O94946 ipomoea bat
44	47	33.8	388	1 CHS8_PPHANT	O22045 pharbitis n
45	47	33.8	388	1 CHS_VIGUN	P51089 vigna ungu

ALIGNMENTS

RESULT 1
CAS2_BOVIN STANDARD; PRT; 222 AA.
ID CAS2_BOVIN Q9TR51;
AC P02663; Q9TR51;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alpha-S2 casein precursor [Contains: Casocidin-1].
GN CSN152.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RX MEDLINE=88188989; PubMed=2833669;
RA Stewart A.F., Bomsing U., Beattie C.W., Shan F., Willis I.M.,
Mackinlay A.G.;
RT "Complete nucleotide sequences of bovine alpha S2- and beta-casein
CDNAs: comparisons with related sequences in other species.";
RL Mol. Biol. Evol. 4:231-241(1987).
RN [2]
RP SEQUENCE OF 16-222 (A ALLELE).
RC TISSUE=Milk;
RX MEDLINE=77185633; PubMed=862906;
RA Brignon G., Ribadeau-Dumas B., Mercier J.-C., Pelissier J.-P.,
Das B.C.;
RT "Complete amino acid sequence of bovine alphaS2-casein.";
RL FEBS Lett. 76:274-279(1977).
RN [3]
RP PARTIAL SEQUENCE (D ALLELE).
RC TISSUE=Milk;
RX MEDLINE=79239837; PubMed=469044;
RA Grosclaude F., Joudrier P., Mahe M.-F.;
RT "A genetic and biochemical analysis of a polymorphism of bovine alpha
S2-casein.";
RL J. Dairy Res. 46:211-213(1979).
RN [4]
RP SEQUENCE OF 165-203, AND CHARACTERIZATION OF CASOCIDIN.
RC TISSUE=Milk;
RX MEDLINE=9600204; PubMed=7556666;
RA Zucht H.-D., Raide M., Adersmann K., Meagert H.-J., Forsmann W.-G.;
RT "Casocidin-I: a casein-alpha 2 derived peptide exhibits antibacterial
activity.";
RL FEBS Lett. 372:185-188(1995).
RN [5]
RP FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CALCIUM PHOSPHATE.
CC - FUNCTION: CASOCIDIN-I INHIBITS THE GROWTH OF ESCHERICHIA COLI AND
STAPHYLOCOCCUS CARNOUS.
CC - SUBCELLULAR LOCATION: Extracellular.
CC - TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC - MASS SPECTROMETRY: MW=4870; METHOD=Electrospray; RANGE=165-203.
CC - POLYMORPHISM: AT LEAST TWO ALLELES EXIST. THE SEQUENCE OF THE A
ALLELE IS SHOWN HERE. THE D ALLELE SEQUENCE DIFFERS FROM THAT
SHOWN IN HAVING A DELETION OF NINE RESIDUES, WHICH MAY BE 49-58.

```

CC 50-59, OR 51-60.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -1- DATABASE: NAME-Protein Spotlight;
CC NOTE: Issue 16 of November 2001.
CC WWW="http://www.expasy.org/spotlight/articles/spl016.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M16644; AAA30479.1; -.
CC PIR; J02008; KAB052.
CC InterPro; IPR001588; Casein.
CC Pfam; PF00363; caseins; 2.
CC PROSITE; PS00306; CASEIN ALPHA BETA; 1.
CC Milk; Phosphorylation; Signal; Repeat; Antibiotic.
CC SIGNAL 1 15
CC FT CHAIN 16 222 ALPHA-S2 CASEIN.
CC FT PEPTIDE 165 203 CASOCIDIN-1.
CC FT REPEAT 76 140
CC FT MOD_RES 23 23 PHOSPHORYLATION.
CC FT MOD_RES 24 24 PHOSPHORYLATION.
CC FT MOD_RES 25 25 PHOSPHORYLATION.
CC FT MOD_RES 31 31 PHOSPHORYLATION.
CC FT MOD_RES 71 71 PHOSPHORYLATION.
CC FT MOD_RES 72 72 PHOSPHORYLATION.
CC FT MOD_RES 73 73 PHOSPHORYLATION.
CC FT MOD_RES 76 76 PHOSPHORYLATION.
CC FT MOD_RES 144 144 PHOSPHORYLATION.
CC FT MOD_RES 146 146 PHOSPHORYLATION.
CC FT CONFLICT 102 102 Q -> E (IN REF. 2).
CC FT SEQUENCE 222 AA; 26019 MW; 81E7408AF1C12E7C CRC64;
SQ
Query Match 100.0%; Score 139; DB 1; Length 222;
Best Local Similarity 100.0%; Pred. No. 2,9e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VYQHOKAMKPMWIOPTKTVIPYRYL 25
DB 199 VYQHOKAMKPMWIOPTKTVIPYRYL 222
RESULT 2
CAS2_CAPHI STANDARD; PRT; 223 AA.
AC P33049;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-S2 casein precursor (Alpha-S2-CN).
GN CSN152.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OC NCB1_TaxID=9925;
OX (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=93216130; PubMed=8462880;
RA Bouniol C.;
RT "Sequence of the goat alpha s2-casein-encoding cDNA.";
RL Gene 125:235-236 (1993).
RN (12)
RP SEQUENCE FROM N.A.
RA Bouniol C., Brignon G., Mahe M.-F., Prinz C.;
RT "Characterization of goat allelic alpha-s2-caseins A and B: further
RT evidence of the phosphorylation code of caseins.";
RL Protein Seq. Data Anal. 5:213-218 (1993).

```

```

RN (3)
RP SEQUENCE FROM N.A. (VARIANT C).
RX MEDLINE=95030556; PubMed=7943951;
RA Bouniol C., Brignon G., Mahe M.-F., Prinz C.;
RT "Biochemical and genetic analysis of variant C of caprine alpha s2-
RT casein (Capra hircus).";
RL Anim. Genet. 25:173-177 (1994).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- POLYMORPHISM: THREE ALLELES OF ALPHA-S2 CASEIN ARE KNOWN: A, B AND
CC C. THE FREQUENCIES OF THE ALLELES IS ESTIMATED TO BE 0.85, 0.04
CC AND 0.11 IN THE FRENCH DAIRY BREEDS 'ALPINE' AND 'SAANEN'.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X65160; CAA46278.1; -.
CC PIR; S74171; AAB32166.1; -.
CC PIR; I46995; I46995.
CC PIR; S33881; JN0547.
CC InterPro; IPR001588; Casein.
CC Pfam; PF00363; caseins; 2.
CC PROSITE; PS00306; CASEIN ALPHA BETA; 1.
CC Milk; Phosphorylation; Signal; Repeat; Polymorphism.
CC SIGNAL 1 15
CC FT CHAIN 16 223 ALPHA-S2 CASEIN.
CC FT REPEAT 77 141
CC FT REPEAT 159 223
CC FT MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 77 77 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 145 145 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 147 147 PHOSPHORYLATION (BY SIMILARITY).
CC FT MOD_RES 179 79 PHOSPHORYLATION (BY SIMILARITY).
CC FT VARIANT 182 182 K -> I (IN VARIANT B).
CC FT VARIANT 182 182 K -> I (IN VARIANT C).
CC FT SEQUENCE 223 AA; 26389 MW; 187DEP42FD688291 CRC64;
SQ
Query Match 82.7%; Score 115; DB 1; Length 223;
Best Local Similarity 84.0%; Pred. No. 9.6e-10;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 VYQHOKAMKPMWIOPTKTVIPYRYL 25
DB 199 VYQHOKAMKPMWIOPTKTVIPYRYL 223
RESULT 3
CAS2_SHEEP STANDARD; PRT; 223 AA.
AC P04654;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-S2 casein precursor.
GN CSN152.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OC NCB1_TaxID=9940;
OX (1)

```

```

RP SEQUENCE FROM N.A.
RA MEDLINE=86104467; PubMed=3002499;
RX Boizard M., Petrisant G.;
RT "Complete sequence of ovine alpha s2-casein messenger RNA.";
RL Biochimie 67:1043-1051(1985).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X03238; CAA26983.1; -.
CC PIR; A25070; KASH82.
CC InterPro; IPR001588; Casein.
CC Pfam; PF00363; caseins; 2.
CC PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
CC MILK; Phosphorylation; Repeat; Signal.
CC SIGNAL 1 15
CC CHAIN 16 223 ALPHA-S2 CASEIN.
CC REPEAT 77 141
CC MOD_RES 159 223
CC MOD_RES 23 223 PHOSPHORYLATION (POTENTIAL).
CC MOD_RES 24 24 PHOSPHORYLATION (POTENTIAL).
CC MOD_RES 25 25 PHOSPHORYLATION (POTENTIAL).
CC MOD_RES 72 72 PHOSPHORYLATION (POTENTIAL).
CC MOD_RES 73 73 PHOSPHORYLATION (POTENTIAL).
CC MOD_RES 74 74 PHOSPHORYLATION (POTENTIAL).
CC VARIANT 64 64 D -> N.
CC SEQUENCE 223 AA; 26332 MW; 67212935E27426D7 CRC64;
SQ
Query Match 82.7%; Score 115; DB 1; Length 223;
Best Local Similarity 84.0%; Pred. No. 9.6e-10;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 VYQHOKAMKPMWIOPTKVIPIYRYL 25
DB 199 VDHQKAMKPMWIOPTKVIPIYRYL 223
-----
RESULT 4
CAS2_PIG STANDARD; PRT; 235 AA.
ID CAS2_PIG
AC P39036;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alpha-S2 casein precursor.
GN CSN152.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sub.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92367960; PubMed=1503276;
RA Alexander L.J., Das Gupta N.A., Beattie C.W.;
RT "The sequence of porcine alpha s2-casein cDNA.";
RL Anim. Genet. 23:365-367(1992).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X54975; CAA38759.1; -.
CC InterPro; IPR001588; Casein.
CC Pfam; PF00363; caseins; 2.
CC PROSITE; PS00306; CASEIN_ALPHA_BETA; 1.
CC MILK; Phosphorylation; Signal.
CC SIGNAL 1 15
CC CHAIN 16 235 ALPHA-S2 CASEIN.
CC MOD_RES 23 23 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 24 24 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 25 25 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 70 70 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 71 71 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 72 72 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 73 73 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 74 74 PHOSPHORYLATION (BY SIMILARITY).
CC SEQUENCE 235 AA; 27570 MW; C903B760D184C14C CRC64;
SQ
Query Match 50.7%; Score 70.5; DB 1; Length 235;
Best Local Similarity 55.6%; Pred. No. 0.0034;
Matches 15; Conservative 4; Mismatches 5; Indels 3; Gaps 1;
QY 1 VYQHOKAMKPMWIOPTKVIPIYRY 24
DB 208 VHQKQAMKPMWIKTNSYQIIPNRY 234
-----
RESULT 5
CAS3_RABIT STANDARD; PRT; 182 AA.
ID CAS3_RABIT
AC P50419;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Alpha-S2B casein precursor.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=New Zealand white; TISSUE=Mammary gland;
MEDLINE=94107245; PubMed=8280077;
RA Dawson S.P., Wilde C.J., Tishe P.J., Mayer R.J.;
RT "Characterization of two novel casein transcripts in rabbit mammary
RT gland.";
RL Biochem. J. 296:777-784(1993).
CC -1- FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
CC -1- SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X76909; CAA54231.1; -.
CC PIR; S39776; S39776.
CC InterPro; IPR001588; Casein.
CC Pfam; PF00363; caseins; 2.
-----

```

```

DR PROSITE; PS00306; CASEIN ALPHA_BETA; 1.
KW Milk; Phosphorylation; signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 182 AA; 21597 MW; F2432C2DA4550FC CRC64;

Query Match
Best Local Similarity 48.2%; Score 67; DB 1; Length 182;
Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 1 VYHQKAMKPWIOPTKVIPYV 23
DB 83 LYQYPTVMDFWTRAEKTAIPFIR 105

RESULT 6
CAS2_CAMDR ID CAS2_CAMDR STANDARD; PRT; 193 AA.
AC 097944;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Alpha-S2 casein precursor.
GN CSNIS2.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Butheraia; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=Somali; TISSUE=tuber;
RX MEDLINE=98291310; PubMed=9627840;
RA Kapeler S., Farah Z., Punan Z.;
RT "Sequence analysis of Camelus dromedarius milk caseins.";
RL J. Dairy Res. 65:209-222(1998).
CC -! FUNCTION: IMPORTANT ROLE IN THE CAPACITY OF MILK TO TRANSPORT
CC CALCIUM PHOSPHATE.
CC -! SUBCELLULAR LOCATION: Extracellular.
CC -! TISSUE SPECIFICITY: Mammary gland; Milk.
CC -! SIMILARITY: BELONGS TO THE ALPHA-CASEIN FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-gsb.ch).
-----
CC CC
CC DR EMBL; AJ012629; CAA10078.1; -.
CC InterPro; IPR001588; Casein.
CC DR Pfam; PF00363; caseins; 2.
CC DR PROSITE; PS00306; CASEIN ALPHA_BETA; FALSE_NEG.
CC KW Milk; Phosphorylation; signal.
CC FT SIGNAL
CC FT CHAIN
CC SQ SEQUENCE 193 AA; 22964 MW; 2843256F8FD2ED13 CRC64;

Query Match
Best Local Similarity 39.6%; Score 55; DB 1; Length 193;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

OY 1 VYHQKAMKPWIOPTKVIPYV 22
DB 93 LHQGIVNPNMDGKTAVPFI 114

RESULT 7
CHSI_LYCSES ID CHSI_LYCSES STANDARD; PRT; 389 AA.
AC P23418;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
```

```

DE 15-JUN-1999 (Rel. 38, Last annotation update)
DR Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1).
GN CHS1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RX [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Cotyledon, Hypocotyl, and Leaf;
RX MEDLINE=91117196; PubMed=1980524;
RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
RT "Molecular genetic analysis of chalcone synthase in Lycopersicon
esculentum and an anthocyanin-deficient mutant.";
RL Mol. Gen. Genet. 224:279-288(1990) .
CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; X55194; CA38980.1; -.
DR InterPro: IPR001099; N-C synthase.
DR Pfam; PF00195; Chal_stil_synth.1.
DR Pfam; PF02797; Chal_stil_synth; 1.
DR ProDom; PD000453; N-C_synthase; 1.
DR PROSITE; PS00441; CHALCONE SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KW Multigene family.
FT ACT_SITE 164 164 BY SIMILARITY
SQ SEQUENCE 389 AA; 42552 MW; 553DC695E5A6AB CRC64;
Query Match 37.4%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 2.9;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 5 QKANKPMIOPKTV 18
DB 111 OKAIKMGOPRSKI 124
RESULT 8
CHS2_LYCES STANDARD; PRT; 389 AA.
ID CHS2_LYCES AC P23419;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1993 (Rel. 38, Last annotation update)
DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RX [1]
RN SEQUENCE FROM N.A.
RC TISSUE=Cotyledon, Hypocotyl, and Leaf;
RX MEDLINE=91117196; PubMed=1980524;

```

RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
 RT "Molecular genetic analysis of chalcone synthase in lycopersicon
 RL esculentum and an anthocyanin-deficient mutant";
 CC Mol. Genet. 224:279-288(1990).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC EMBL: X55195; CAA38981.1; -;
 CC InterPro: IPR001099; N-C_synthase.
 CC Pfam: PF00195; Chal_stil_synth; 1.
 CC Pfam: PF02797; Chal_stil_synth; 1.
 CC ProDom: PD000453; N-C_synthase; 1.
 CC PROSITE: PS000441; CHALCONE_SYNTH; 1.
 CC Flavonoid biosynthesis; Transferase; Acyltransferase;
 CC Multigene family.
 CC ACT SITE 164 164
 CC FT ACT SITE 164 164 BY SIMILARITY.
 CC SQ SEQUENCE 389 AA; 42730 MW; F92E46B3B5FC32F CRC64;

Query Match 37.4%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 2.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAKMPWIOPKTKV 18
 Db 111 OKAIKMGQPKSKI 124

RESULT 9
 CHS2_SOLTV STANDARD; PRT; 389 AA.
 AC QM3188;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
 GN CHS2.
 OS Solanum tuberosum (Potato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
 OC NCBI_TaxID=4113;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Red Pontiac;
 RX MEDLINE=97141614; PubMed=8987872;
 RA Jeon J.-H., Kim H.S., Choi K.H., Young Y.H., Young H., Byun S.-M.;
 RT "Cloning and characterization of one member of the chalcone synthase
 RT gene family from Solanum tuberosum L.";
 RL Biosci. Biotechnol. Biochem. 60:1907-1910(1996).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY

CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC EMBL: U47738; AAB05239.1; -;
 CC PIR: JCS136; JCS136.
 CC InterPro: IPR001099; N-C_synthase.
 CC Pfam: PF00195; Chal_stil_synth; 1.
 CC Pfam: PF02797; Chal_stil_synth; 1.
 CC ProDom: PD000453; N-C_synthase; 1.
 CC PROSITE: PS000441; CHALCONE_SYNTH; 1.
 CC Flavonoid biosynthesis; Transferase; Acyltransferase;
 CC Multigene family.
 CC ACT SITE 164 164
 CC FT ACT SITE 164 164 BY SIMILARITY.
 CC SQ SEQUENCE 389 AA; 42476 MW; 41618F944958603 CRC64;

Query Match 37.4%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 2.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAKMPWIOPKTKV 18
 Db 111 OKAIKMGQPKSKI 124

RESULT 10
 CHSA_PETHY STANDARD; PRT; 389 AA.
 AC P08894;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Chalcone synthase A (EC 2.3.1.74) (Naringenin-chalcone synthase A).
 GN CHSA.
 OS Petunia hybrida (Petunia).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
 OC NCBI_TaxID=4102;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Violet 30;
 RX MEDLINE=86286540; PubMed=3016642;
 RA Koes R.E., Spelt C.E., Reif H.J., van den Elzen P.J.M., Velthamp E.,
 RA Mol J.N.M.;
 RT "Floral tissue of Petunia hybrida (V30) expresses only one member of
 RT the chalcone synthase multigene family.";
 RL Nucleic Acids Res. 14:5229-5239(1986).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Violet 30; TISSUE=leaf;
 RX MEDLINE=90034197; PubMed=2806915;
 RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;
 RT "Cloning and molecular characterization of the chalcone synthase
 RT multigene family of Petunia hybrida.";
 RL Gene 81:245-257(1989).
 CC -1- FUNCTION: The primary product of this enzyme is 4,2',4',6'-
 CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
 CC which can under specific conditions spontaneously isomerize into
 CC naringenin.
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
 CC naringenin-chalcone + 3 CO(2).
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
 CC OF WHICH ARE BRIGHTLY COLORED.
 CC -1- TISSUE SPECIFICITY: MAJOR EXPRESSED MEMBER OF THE GENE FAMILY IN

```
CC VARIOUS FLORAL TISSUES AND IN SEEDLINGS TREATED WITH UV LIGHT. IT
CC IS RELATIVELY LOW EXPRESSED IN TISSUE CULTURE MATERIAL.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
-----
DR EMBL; X04080; CAA27718.1; -.
DR EMBL; X14591; CAA32731.1; -.
DR PIR; A23643; SYPON.
DR PIR; J50308; SYPOCA.
DR InterPro; IPR001099; N-C synthase.
DR Pfam; PF00195; Chal_stil_synthc; 1.
DR Pfam; PF02797; Chal_stil_synthc; 1.
DR ProDom; PD000453; N-C synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KW Multigene family.
FT ACT SITE 164
FT CONFLICT 128
SQ SEQUENCE 389 AA; 42525 MW; 0221B8DA1E96EAF6 CRC64;
BY SIMILARITY.
V-> F (IN REF. 2).
Query Match 37.4%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. NO. 2.9;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 5 OKAKKWTQPKTV 18
DB 111 OKAIKEWGQPSKI 124
|||:|||:|
|||:|||:|
RESULT 11
CHSA_SOLTU STANDARD; PRG; 389 AA.
ID CHSA_SOLTU Q41436;
AC Q41436;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chalcone synthase 1A (EC 2.3.1.74) (Naringenin-chalcone synthase 1A).
GN CHS1A.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxId=4113;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Red Pontiac;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Young Y.-H., Young H., Byun S.-M.;
RT "Characterization of two members of the chalcone synthase gene family
RT from Solanum tuberosum L.";
RL (In) Plant Gene Register PGR96-027.
CC -1- FUNCTION: The primary product of this enzyme is 4',2'',4'',6''-
CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES. MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/  
or send an email to license@lsb-sib.ch).  
-----  
DR EMBL; U47739; AAB67734.1; '-  
DR InterPro; IPR001099; N-C_synthase.  
DR Pfam; PF00195; Chal_stil_synC; 1.  
DR Pfam; PF02797; Chal_stil_synC; 1.  
DR ProDom; PD000453; N-C_synthase; 1.  
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.  
KW Flavonoid biosynthesis; Transferase; Acyltransferase;  
KW Multigene family.  
FT ACT SITE 164  
SQ SEQUENCE 389 AA; 42562 MW; B18ID9C6B9170A34 CRC64;  
  
Query Match 37.4%; Score 52; DB 1; Length 389;  
Best local Similarity 64.3%; Pred. No. 2.9;  
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
  
OY 5 OKAKPWIQPTKY 18  
|||:|||||:  
Db 111 QKAKEWGPQPSKI 124  
  
RESULT 12  
CHSB_SOLTU STANDARD; PRT; 389 AA.  
AC Q43163;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Chalcone synthase 1B (EC 2.3.1.74) (Naringenin-chalcone synthase 1B).  
GN CHS1B.  
OS Solanum tuberosum (potato); Streptophyta; Embryophyta; Tracheophyta;  
OC Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamids; Solanales; Solanales; Solanales; Solanaceae; Solanum.  
CX NCBI_TaxID=4113;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=cv. Red Pontiac;  
RC Jeon J.-H., Kim H.-S., Choi K.-H., Joung Y.-H., Joung H., Byun S.-M.;  
RT "Characterization of two members of the chalcone synthase gene family  
from Solanum tuberosum L.";  
RL (In) Plant Gene Register PCR96-027.  
CC -1- FUNCTION: The primary product of this enzyme is 4,2',4'',6'-  
tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)  
which can under specific conditions spontaneously isomerize into  
naringenin.  
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +  
naringenin-chalcone + 3 CO(2).  
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF  
FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY  
OF WHICH ARE BRIGHLY COLORED.  
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.lsb-sib.ch/announcement/  
CC or send an email to license@lsb-sib.ch).  
-----  
DR EMBL; U47740; AAB67735.1; '-  
DR InterPro; IPR001099; N-C_synthase.  
DR Pfam; PF00195; Chal_stil_synC; 1.  
DR Pfam; PF02797; Chal_stil_synC; 1.  
DR ProDom; PD000453; N-C_synthase; 1.  
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.  
KW Flavonoid biosynthesis; Transferase; Acyltransferase;  
KW Multigene family.  
FT ACT SITE 164  
SQ SEQUENCE 389 AA; 42548 MW; E7333A1A34BD1CC CRC64;
```



```

OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Petunia.
OX NCBI_TaxId=4102;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Violet 30; TISSUE=leaf;
RX MEDLINE=90034197; PubMed=2806915;
RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;
RT "Cloning and molecular characterization of the chalcone synthase
RT multigene family of Petunia hybrida.";
RL Gene 81:245-257(1989).
RN [2]
RP SEQUENCE OF 71-389 FROM N.A.
RC STRAIN=cv. White 137; TISSUE=Anther;
RA van Tunen A.J.;
RL Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: The primary product of this enzyme is 4',2',4',6'-
CC tetrahydroxychalcone (also termed naringenin-chalcone or chalcone)
CC which can under specific conditions spontaneously isomerize into
CC naringenin.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14597; CAA32737.1; -.
DR EMBL; X14599; CAA32739.1; -.
DR PIR; D72821; SYPJGJ.
DR PIR; S18136; S18136.
DR InterPro; IPR001099; N_C_synthase.
DR Pfam; PF00195; Chal_stil_synth; 1.
DR Pfam; PF02797; Chal_stil_synthC; 1.
DR ProDom; PD000453; N-C synthase; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KW Multigene family.
FT ACT_SITE 164 BY SIMILARITY.
FT CONFLICT 75 E -> V (IN REF. 2).
SQ SEQUENCE 389 AA; 42558 MW; F2B3CDD82E6FDE7D CRC64;

```

Query Match 37.4%; Score 52; DB 1; Length 389;
 Best Local Similarity 64.3%; Pred. No. 2.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 QXAKKPIQPKTKV 18
 |||:| |||:|
 Db 111 QXAIKMGQPKSKI 124

Search completed: July 30, 2003, 16:24:29
 Job time : 9.18452 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 15.9226 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-4

Perfect score: 139
Sequence: 1 VYQHOKAMKRWIQPKTKVPIVRYL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR 76: *
2: PIR1: *
3: PIR2: *
4: PIR3: *
5: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	139	100.0	232	1 KABS02	alpha-s2-casein pr
2	115	82.7	223	1 KABS02	alpha-s2-casein pr
3	115	82.7	223	2 JN0547	alpha-s2-casein pr
4	115	82.7	223	2 I46995	alpha-s2-casein C
5	70.5	50.7	235	2 A48383	alpha-S2-casein -
6	67	48.2	182	2 S39776	probable DNA damag
7	53	38.1	935	2 T39389	alpha-s2-casein fo
8	52	37.4	319	2 S18136	naringenin-chalcon
9	52	37.4	340	2 T10957	naringenin-chalcon
10	52	37.4	389	1 SYPOCJ	naringenin-chalcon
11	52	37.4	389	1 SYPOCJ	naringenin-chalcon
12	52	37.4	389	1 SYPOCJ	naringenin-chalcon
13	52	37.4	389	1 SYPOCJ	naringenin-chalcon
14	52	37.4	389	2 T07789	naringenin-chalcon
15	52	37.4	389	2 T07789	naringenin-chalcon
16	51	36.7	180	2 S39775	alpha-s2-casein fo
17	50.5	36.3	471	2 F71618	adenylsuccinate 1
18	50	36.0	311	2 T15997	hypothetical prote
19	50	36.0	1877	2 T40550	probable phosphati
20	49.5	35.6	325	2 C64000	hypothetical prote
21	49.5	35.6	325	2 C64000	hypothetical prote
22	49	35.3	188	2 J04680	probable zinc meta
23	49	35.3	207	2 J04679	vascular endotheli
24	49	35.3	342	2 AD2032	hypothetical prote
25	49	35.3	398	2 S16275	naringenin-chalcon
26	49	35.3	400	1 SYZMCC	naringenin-chalcon
27	48	34.5	392	2 S16206	stilbene synthase
28	48	34.5	392	2 S11044	stilbene synthase
29	48	34.5	392	2 S53313	stilbene synthase

30	48	34.5	392	2 S53314	stilbene synthase
31	48	34.5	418	2 F72370	phob-related prote
32	47	33.8	53	2 S41957	naringenin-chalcon
33	47	33.8	128	2 S33611	naringenin-chalcon
34	47	33.8	158	2 B66169	hypothetical prote
35	47	33.8	285	2 S44369	naringenin-chalcon
36	47	33.8	331	2 J01071	naringenin-chalcon
37	47	33.8	370	2 S44368	naringenin-chalcon
38	47	33.8	383	2 S35165	naringenin-chalcon
39	47	33.8	388	1 SYGYCN	naringenin-chalcon
40	47	33.8	388	1 SYGYC1	naringenin-chalcon
41	47	33.8	388	1 SYGYC3	naringenin-chalcon
42	47	33.8	388	2 S37098	naringenin-chalcon
43	47	33.8	388	2 J02249	naringenin-chalcon
44	47	33.8	388	2 J02259	naringenin-chalcon
45	47	33.8	388	2 S60472	naringenin-chalcon

ALIGNMENTS

RESULT 1
KABS02
alpha-s2-casein precursor - bovine
C/Species: Bos primigenius taurus (cattle)
C/Date: 01-Sep-1981 #sequence, revision 03-Feb-1994 #ext_change 22-Jun-1999
C/Accession: J02008; A29087; A91438; S66626; A03107
R/Groenen, M.A.M.; Dijkhof, R.J.M.; Verstege, A.J.M.; van der Poel, J.J.
Gene 123, 187-193, 1993
A/Title: The complete sequence of the gene encoding bovine alphas2-casein.
A/Reference number: J02008; MUID:93154583; PMID:8428658
A/Accession: J02008
A/Status: translation not shown
A/Molecule type: DNA
A/Residues: 1-222 <GRO>
A/Cross-references: GB:M94327
R/Stewart, A.F.; Bonsing, J.; Beattie, C.W.; Shah, F.; Willis, I.M.; Mackinlay, A.G.
Mol. Biol. Evol. 4, 231-241, 1987
A/Title: Complete nucleotide sequences of bovine alpha-s2- and beta-casein cDNAs: compa
A/Reference number: A93062; MUID:88188989; PMID:2833669
A/Accession: A29087
A/Status: translation not shown
A/Molecule type: mRNA
A/Residues: 1-222 <STB>
A/Cross-references: GB:M16644; NID:G162928; PIDN:AA30479.1; PID:G162929
R/Brignon, G.; Ribadeau Dumas, B.; Mercier, J.C.; Pelissier, J.P.; Dae, B.C.
FEBS Lett. 76, 274-279, 1977
A/Title: Complete amino acid sequence of bovine alpha-S2-casein.
A/Reference number: A91438; MUID:77185633; PMID:862906
A/Contents: A allele
A/Accession: A91438
A/Molecule type: protein
A/Residues: 16-101, 'BE', 104-222 <BRI>
A/Note: Four fractions, previously designated s2, s3, s4, and s6, appear to have the sa
these
J. Dairy Res. 46, 211-213, 1979
A/Title: A genetic and biochemical analysis of a polymorphism of bovine alpha-s2-casein
A/Reference number: A92771; MUID:79239837; PMID:469044
A/Contents: annotation; D allele
A/Note: the sequence of the D allele has a deletion of nine residues, which may be 49-5
R/Zucht, H.D.; Raide, M.; Adernann, K.; Maeger, H.J.; Forssmann, W.G.
FEBS Lett. 372, 185-188, 1995
A/Title: Casocidin-I, a casein-alpha(s2) derived peptide exhibits antibacterial activit
A/Reference number: S66626; MUID:9600204; PMID:755666
A/Accession: S66626
A/Molecule type: protein
A/Residues: 165-203 <ZUC>
C/Comment: The sequence of the A allele is shown.
C/Genetics:
A/Gene: alphas2a
A/Map position: 6
A/Intons: 17/3; 26/3; 33/3; 47/3; 56/3; 65/3; 74/3; 82/3; 97/3; 138/3; 147/3; 156/3; 1

C/Superfamily: alpha-s2-casein
 C/Keywords: mammary gland; milk; phosphoprotein
 F/1-15/Domain: signal sequence #status predicted <SIG>
 F/16-223/Product: alpha-s2-casein #status experimental <MAT>
 F/23,24,25,31,71,72,73,76,144,146,156/Binding site: phosphate (Ser) (covalent) #status F

Query Match 100.0%; Score 139; DB 1; Length 222;
 Best Local Similarity 100.0%; Pred. No. 46-13;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKTKVIPYRYL 25
 DB 199 VYQHOKAMKPMIOPKTKVIPYRYL 222

RESULT 2

KASHS2
 alpha-s2-casein precursor - sheep
 C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C/Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 22-Jun-1999
 C/Accession: A25070; S17856
 R/Bouma, M.; Petrisant, G.
 Biochimie 67, 1043-1051, 1985
 A/Title: Complete sequence of ovine alpha-s2-casein messenger RNA.
 A/Reference number: A25070; MUID:86104467; PMID:3002499
 A/Accession: A25070
 A/Molecule type: mRNA
 A/Residues: 1-223 <BO1>

A/Cross-references: GB:X03238; NID:gl238; PIDN:CAA26983.1; PID:9732894
 A/Note: 64-Aam was also found
 R/Bouma, M.; Hue, D.; Bouniol, C.; Mercier, J.C.; Gayer, P.
 Eur. J. Biochem. 201, 633-641, 1991
 A/Title: Multiple mRNA species code for two non-allelic forms of ovine alpha-s2-casein.
 A/Reference number: S17856; MUID:92037619; PMID:1935959
 A/Accession: S17856
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-93, 'T', '95-223 <BO2>
 C/Superfamily: alpha-s2-casein
 C/Keywords: mammary gland; milk; phosphoprotein
 F/1-15/Domain: signal sequence #status predicted <SIG>
 F/16-223/Product: alpha-s2-casein #status predicted <KA2>
 F/23,24,25,32,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status F/23,24,25,31,71,72,73,76,144,146,156/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 82.7%; Score 115; DB 1; Length 223;
 Best Local Similarity 84.0%; Pred. No. 1.5e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKTKVIPYRYL 25
 DB 199 VYQHOKAMKPMIOPKTKVIPYRYL 223

RESULT 3

JN0547
 alpha-s2-casein precursor - goat
 C/Species: Capra aegagrus hircus (domestic goat)
 C/Date: 31-Dec-1993 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C/Accession: S33881; S33880; JN0547; S20620
 R/Bouma, M.; Brignon, G.; Mahe, M.F.; Prinz, C.
 Protein Seq. Data Anal. 5, 213-218, 1993
 A/Title: Characterization of goat allelic alpha-s2-caseins A and B: further evidence of

A/Accession: S33881
 A/Status: preliminary
 A/Molecule type: protein
 A/Residues: 1-223 <BOU>
 A/Note: allele B
 A/Accession: S33880

A/Molecule type: protein
 A/Residues: 16-78, 'E', '80-223 <BO2>
 A/Note: sequence deduced from compositional analysis of peptides

R/Bouma, M.
 Gene 125, 235-236, 1993
 A/Title: Sequence of the goat alpha-s2-casein-encoding cDNA.
 A/Reference number: JN0547; MUID:93216130; PMID:8462880
 A/Accession: JN0547

A/Molecule type: mRNA
 A/Residues: 1-78, 'E', '80-223 <BO3>
 A/Cross-references: EMBL:X65160; NID:9955; PIDN:CAA46278.1; PID:9956
 A/Note: allele A
 C/Superfamily: alpha-s2-casein

C/Keywords: mammary gland; milk; phosphoprotein
 F/1-15/Domain: signal sequence #status predicted <SIG>
 F/16-223/Product: alpha-s2-casein #status experimental <MAT>
 F/23,24,25,55,72,73,74,77,145,147,159/Binding site: phosphate (Ser) (covalent) #status F/23,24,25,31,71,72,73,76,144,146,156/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 82.7%; Score 115; DB 2; Length 223;
 Best Local Similarity 84.0%; Pred. No. 1.5e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKTKVIPYRYL 25
 DB 199 VYQHOKAMKPMIOPKTKVIPYRYL 223

RESULT 4

I46995
 alpha s2-casein C - goat
 C/Species: Capra aegagrus hircus (domestic goat)
 C/Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 13-Aug-1999
 C/Accession: I46995
 R/Bouma, M.; Brignon, G.; Mahe, M.F.; Prinz, C.
 Anim. Genet. 25, 173-177, 1994
 A/Title: Biochemical and genetic analysis of variant C of caprine alpha s2-casein (Capra
 A/Reference number: I46995; MUID:95030556; PMID:7943651
 A/Accession: I46995
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-223 <BOU>
 A/Cross-references: GB:S74171; NID:9707033; PIDN:AA83216.1; PID:9707034
 C/Superfamily: alpha-s2-casein

Query Match 82.7%; Score 115; DB 2; Length 223;
 Best Local Similarity 84.0%; Pred. No. 1.5e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMIOPKTKVIPYRYL 25
 DB 199 VYQHOKAMKPMIOPKTKVIPYRYL 223

RESULT 5

A48383
 alpha s2-casein - pig
 C/Species: Sus scrofa domestica (domestic pig)
 C/Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 03-May-1996
 C/Accession: A48383
 R/Alexander, L.J.; Das Gupta, N.A.; Beattie, C.W.
 Anim. Genet. 23, 365-367, 1992
 A/Title: The sequence of porcine alpha s2-casein cDNA.
 A/Reference number: A48383; MUID:92367960; PMID:1503276
 A/Accession: A48383
 A/Status: preliminary
 A/Molecule type: nucleic acid
 A/Residues: 1-225 <ALB>
 A/Experimental source: mammary gland
 A/Note: sequence inconsistent with the nucleotide translation
 A/Note: sequence extracted from NCBI backbone (NCBIN:110884, NCBI:110885)
 C/Superfamily: alpha-s2-casein

Query Match 50.7%; Score 70.5; DB 2; Length 235;
 Best Local Similarity 55.6%; Pred. No. 0.0062;
 Matches 15; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

QY 1 VYOHOKAMKPMWIOPTKV---KVIPYRVY 24
 DB 208 VHQKOKAMKPMWIKTKNSYQIIPMLAY 234

RESULT 6

alpha-S2-casein form b precursor - rabbit
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C:Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 13-Aug-1999
 C:Accession: S39776
 R:Daewon, S.P.; Wilde, C.J.; Tighe, P.J.; Mayer, R.J.
 Biochem. J. 296, 777-784, 1993
 A:Title: Characterization of two novel casein transcripts in rabbit mammary gland.
 A:Reference number: S39775, PMID:94107245, PMID:8280077
 A:Accession: S39776
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-182 <DAM>
 A:Cross-references: EMBL:X76909; NID:g439527; PIDN:CAA54231.1; PID:g439528
 C:Superfamily: alpha-S2-casein

Query Match 48.2%; Score 67; DB 2; Length 182;
 Best Local Similarity 43.5%; Pred. No. 0.016;
 Matches 10; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 VYOHOKAMKPMWIOPTKVIPYRV 23
 DB 83 LYQYPTVMPDPMTRAKTKALPFR 105

RESULT 7

probable DNA damage repair protein SPBC1347.01c SPBC215.16c - fission yeast (Schizosacch
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Sep-2000
 C:Accession: T39389, T39906
 R:Wood, V.; Rajandream, M.A.; Barrrell, B.G.; Rieger, M.
 submitted to the EMBL Data Library, February 1999
 A:Reference number: Z21851
 A:Accession: T39389
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-935 <WCO>
 A:Cross-references: EMBL:AL035548; PIDN:CA837432.1; GSPDB:GN00067; SPDB:SPBC1347.01c
 R:Lyne, M.; Rajandream, M.A.; Barrrell, B.G.; Rieger, M.
 submitted to the EMBL Data Library, November 1998
 A:Reference number: Z21889
 A:Accession: T39906
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 731-935 <LVN>
 A:Cross-references: EMBL:AL035354; PIDN:CAA22130.1; GSPDB:GN00067; SPDB:SPBC215.16c
 C:Genetics:
 A:Gene: SPDB:SPBC1347.01c
 A:Map position: 2

Query Match 38.1%; Score 53; DB 2; Length 935;
 Best Local Similarity 37.0%; Pred. No. 10;
 Matches 10; Conservative 8; Mismatches 5; Indels 4; Gaps 2;

QY 2 YOHOKAMKPMWIO---PKRKVIPYRVY 24
 DB 119 WKHQKVKPMWIVDCIKOKKILPIMVY 145

RESULT 8

S18136
 naringenin-chalcone synthase (EC 2.3.1.74) - garden petunia
 C:Species: Petunia x hybrida (garden petunia)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 05-May-2000

C:Accession: S18136
 R:van Tunen, A.J.
 submitted to the EMBL Data Library, March 1989
 A:Reference number: S18136
 A:Accession: S18136
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-319 <TUN>
 A:Cross-references: EMBL:X14599; NID:g20537; PIDN:CAA32739.1; PID:g20538
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 2; Length 319;
 Best Local Similarity 64.3%; Pred. No. 4.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OQAMKPMWIOPTKV 18
 DB 41 OQAIKEMGQPKSKI 54

RESULT 9

naringenin-chalcone synthase (EC 2.3.1.74) CHS-FL1 - common morning-glory
 C:Species: Ipomoea purpurea (common morning-glory)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 05-May-2000
 C:Accession: T10957
 R:Rausher, M.D.; Tiffin, P.L.; Miller, R.E.
 submitted to the EMBL Data Library, January 1997
 A:Description: Regulation of anthocyanin gene expression in Ipomoea purpurea.
 A:Reference number: Z17229
 A:Accession: T10957
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-340 <RAU>
 A:Cross-references: EMBL:U74082; NID:g1658119; PID:g1658120
 A:Experimental source: flower bud; WMAA genotype (dark flowered)
 C:Genetics:
 A:Gene: CHS-FL1
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A

Query Match 37.4%; Score 52; DB 2; Length 340;
 Best Local Similarity 64.3%; Pred. No. 5.1;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OQAMKPMWIOPTKV 18
 DB 111 OQAIKEMGQPKSKI 124

RESULT 10

SYRPTJ
 naringenin-chalcone synthase (EC 2.3.1.74) J - garden petunia
 N:Alternate names: chalcone synthase
 C:Species: Petunia x hybrida (garden petunia)
 C:Date: 30-Sep-1991 #sequence_revision 17-Mar-2000 #text_change 05-May-2000
 C:Accession: D72821; J50309
 R:Koes, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.
 Gene 81, 245-257, 1989
 A:Title: Cloning and molecular characterization of the chalcone synthase multigene fami
 A:Reference number: J50308; PMID:90034197; PMID:2806915
 A:Accession: D72821

A:Molecule type: DNA
 A:Residues: 1-389 <KOE>
 A:Cross-references: EMBL:X14597; NID:g20535; PIDN:CAA32737.1; PID:g20536
 A:Experimental source: strain Violet 30, leaf
 A:Accession: J50309

A:Molecule type: DNA
 A:Residues: 1-50, 'D', '52-74', 'V', '76-228', 'I', '230-297', 'L', '299-389 <KO2>
 A>Note: the sequence is revised in Genbank entry PHCHSJ release 114, (PIDN:CAA32737.1)
 C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavon

A:Gene: chsJ
A:Map position: V
A:Note: chsJ is expressed in various floral tissues and UV illuminated seedlings
C:Superfamily: chalcone synthase
C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 5.8;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18
DB 111 OKAIKMGQPKSKI 124

RESULT 11

SYRJCN
naringenin-chalcone synthase (EC 2.3.1.74) R - garden petunia

N:Alternate names: chalcone synthase
C:Species: Petunia x hybrida (garden petunia)

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000

C:Accession: A23643

R:Koev, R.E.; Spelt, C.E.; Reif, H.U.; van den Elzen, P.J.M.; Veltkamp, E.; Mol, J.N.M.

Nucleic Acids Res. 14, 5225-5239, 1986

A:Title: Floral tissue of Petunia hybrida (V30) expresses only one member of the chalcone synthase gene family

A:Reference number: A23643; MUID:86286540; PMID:3016642

A:Accession: A23643

A:Molecule type: mRNA

A:Residues: 1389 <KOE>

A:Cross-references: GB:X04080; NID:g20541; PIDN:CAA27718.1; PID:g20542

A:Experimental source: strain Violet 30, flowers

C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoids

C:Genetics:

A:Gene: chsR

A:Note: expressed in floral tissue

C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 5.8;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18
DB 111 OKAIKMGQPKSKI 124

RESULT 12
SYRJCA
naringenin-chalcone synthase (EC 2.3.1.74) A - garden petunia

N:Alternate names: chalcone synthase

C:Species: Petunia x hybrida (garden petunia)

C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 05-May-2000

C:Accession: J03038

R:Koev, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.

Gene 81, 245-257, 1989

A:Title: Cloning and molecular characterization of the chalcone synthase multigene family

A:Reference number: J03038; MUID:90034197; PMID:2806915

A:Accession: J03038

A:Molecule type: DNA

A:Residues: 1389 <KOE>

A:Cross-references: GB:X14591; NID:g20524; PIDN:CAA32731.1; PID:g20525

A:Experimental source: strain Violet 30, leaf

C:Comment: This enzyme plays a central role in the biosynthesis of all classes of flavonoids

C:Genetics:

A:Gene: chsA

A:Map position: V

A:Note: chsA is the major expressed member of the gene family in various floral tissues & C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 1; Length 389;
Best Local Similarity 64.3%; Pred. No. 5.8;

Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18
DB 111 OKAIKMGQPKSKI 124

RESULT 13

SYRJCN
naringenin-chalcone synthase (EC 2.3.1.74) 2 - potato

N:Alternate names: chalcone synthase; CHS

C:Species: Solanum tuberosum (potato)

C>Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 05-May-2000

C:Accession: J05136; PC4239

R:Jeon, J.H.; Kim, H.S.; Choi, K.H.; Jeong, Y.H.; Jeong, H.; Byun, S.M.

Biochem. Biotechnol. Biochem. 60, 1907-1910, 1996

A:Title: Cloning and characterization of one member of the chalcone synthase gene family

A:Reference number: J05136; MUID:97141614; PMID:987872

A:Accession: J05136

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1389 <JEO>

A:Cross-references: GB:U47738; NID:g1470059; PIDN:AB05239.1; PID:g1470060

A:Accession: PC4239

A:Status: preliminary

A:Molecule type: protein

A:Residues: 158-165;367-373 <JEO2>

C:Comment: This enzyme is important in the biosynthesis of all classes of flavonoids in C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A

Query Match 37.4%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 5.8;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18
DB 111 OKAIKMGQPKSKI 124

RESULT 14
T07799
naringenin-chalcone synthase (EC 2.3.1.74) - common morning-glory

N:Alternate names: chalcone synthase

C:Species: Ipomoea purpurea (common morning-glory)

C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 21-Jul-2000

C:Accession: T07799

R:Fukada-Tanaka, S.; Hoshino, A.; Hisatomi, Y.; Habu, Y.; Hasebe, M.; Iida, S.

Plant Cell Physiol. 38, 754-758, 1997

A:Title: Identification of new chalcone synthase genes for flower pigmentation in the JA

A:Reference number: Z16140; MUID:97393496; PMID:9249990

A:Accession: T07799

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1389 <FUK>

A:Cross-references: EMBL:AB001827; NID:g2329836; PIDN:BAA21789.1; PID:g2329837

C:Genetics:

A:Gene: CHD-E

C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 2; Length 389;
Best Local Similarity 64.3%; Pred. No. 5.8;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPKTKV 18
DB 111 OKAIKMGQPKSKI 124

RESULT 15

SYSKCD
naringenin-chalcone synthase (EC 2.3.1.74) - garden snapdragon

N:Alternate names: chalcone synthase
 C:Species: Antirrhinum majus (garden snapdragon)
 C>Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #ext_change 05-May-2000
 C:Accession: S07312; A33217
 R:Sommer, H.; Saedler, H.
 Mol. Gen. Genet. 202, 429-434, 1986
 A>Title: Structure of the chalcone synthase gene of Antirrhinum majus.
 A:Reference number: S07312
 A:Accession: S07312
 A:Molecule type: DNA
 A:Residues: 1-390 <SOM>
 A:Cross-references: EMBL:X03710; NID:g16015; PID:CAA27338.1; PID:g16016
 A:Accession: A33217
 A>Status: translation not shown
 A:Molecule type: mRNA
 A:Residues: 1-390 <SOM2>
 C:Genetics:
 A:Gene: chs
 A:Introns: 60/1, 162/3
 C:Superfamily: chalcone synthase
 C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 37.4%; Score 52; DB 1; Length 390;
 Best Local Similarity 64.3%; Pred. No. 5.9;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 OKAMKPMIOPTKV 18
 |||:| | |||:
 Db 111 OKAIKEMGQPKSKI 124

Search completed: July 30, 2003, 16:31:15
 Job time : 15.9226 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 49.256 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-4

Perfect score: 139

Sequence: 1 VYQHOKAMKPMWTPKTKVTPYRYL 25

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2002.DAT:*
- 24: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	139	100.0	25	13	AA20055
2	139	100.0	25	21	AAV88340
3	139	100.0	31	18	AAW32219
4	139	100.0	31	13	AAE17466
5	139	100.0	232	18	AAW32220
6	139	100.0	232	23	AAE17468
7	135	97.1	24	15	AA60481
8	135	97.1	24	21	AAV88339
9	128.5	92.4	26	20	AAW93885

10	119	85.6	21	21	AAV88338
11	115	82.7	223	23	AAE17469
12	115	82.7	223	23	AAE17470
13	115	82.7	223	23	AAE17471
14	115	82.7	223	23	AAE17472
15	115	82.7	223	23	AAE17473
16	105	75.5	19	18	AAW32218
17	105	75.5	19	18	AAE17465
18	92	66.2	16	21	AAV88337
19	70.5	50.7	234	23	AAE17474
20	69	49.6	13	23	AAE17464
21	67	48.2	178	23	AAE17472
22	64	46.0	12	23	AAE17463
23	57	41.0	11	18	AAW32217
24	57	41.0	11	23	AAE17462
25	53	38.1	9	17	AAW04282
26	52	37.4	10	18	AAW32216
27	52	37.4	10	23	AAE17461
28	52	37.4	360	22	AAE60169
29	52	37.4	389	17	AAE96125
30	52	37.4	389	22	AAE74019
31	52	37.4	389	24	ABG73901
32	51	36.7	394	24	ABE60374
33	51	36.7	399	24	ABE60375
34	50.5	36.3	471	21	AAE18201
35	50	36.0	291	22	AAE69092
36	49	35.3	8	17	AAW04278
37	49	35.3	133	17	AAW04828
38	49	35.3	133	20	AAW80492
39	49	35.3	133	23	AAU83406
40	49	35.3	160	20	AAW86208
41	49	35.3	167	20	AAW86234
42	49	35.3	188	17	AAW04826
43	49	35.3	188	17	AAW00864
44	49	35.3	188	20	AAW86201
45	49	35.3	188	20	AAW80490

ALIGNMENTS

RESULT 1
AA20055
AA20055 standard; peptide; 25 AA.
AC AAR20055;
XX
DT 25-MAR-2003 (updated)
DT 26-MAR-1992 (first entry)
XX
DE Casein peptide.
XX
KW Casein; platelet; aggregation; thrombosis; collagen.
XX
OS Synthetic.
XX
PN JP03255095-A.
PD 13-NOV-1991.
XX
PF 02-MAR-1990; 90JP-0052554.
XX
PR 02-MAR-1990; 90JP-0052554.
XX
PA (KANE) KANEBO LTD.
XX
DR WPI; 1992-002669/01.
XX
PT Casein peptide(s) for treating thrombosis - as inhibitors of
PT platelet aggregation caused by adenosine-5'-di-phosphate and
PT collagen and as biochemical reagents
XX
PS Claim 1; Page 1; 8pp; Japanese.

Internal peptide f
Alpha-S2 casein pr
Goat alpha-S2 case
Capra hircus alpha
Sheep alpha-S2 cas
Alpha-S2 casein pr
AAW32218
Alpha-S2 casein pr
Bovine alpha-S2 ca
Internal peptide f
Pig alpha-S2 casei
Bovine alpha-S2 ca
Rabbit alpha-S2 ca
Bovine alpha-S2 ca
Alpha-S2 casein pr
Bovine alpha-S2 ca
Milk derived anti-
Alpha-S2 casein pr
Bovine alpha-S2 ca
Bovine alpha-S2 ca
Pertuna chalcone s
Chalcone synthase.
Cyclamen chalcone
Pertuna chalcone s
Humulus lupulus ch
Humulus lupulus ch
Plasmodium falcipar
Mouse GN protein s
Milk derived anti-
Vascular endotheli
Murine vascular en
Vascular endotheli
Human VEGF-B trunc
Human VEGF-B full
Heart vascular end
Murine VPR167. Mu
Human vascular end
Murine vascular en

XX Ieu25 may be omitted. The peptide and its salts have inhibiting
CC activity against platelet aggregation caused by adenosine-5'-
CC diphosphate and by collagen. They are useful for prevention and
CC treatment of thrombosis and are also useful as a biochemical reagent.
CC (Updated on 25-MAR-2003 to correct PA field.)

XX SQ Sequence 25 AA;

Query Match 100.0%; Score 139; DB 13; Length 25;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMWIOPKTKVIPYRYL 25
Db 1 VYQHOKAMKPMWIOPKTKVIPYRYL 25

RESULT 2
AAV88340

ID AAV88340 standard; peptide; 25 AA.

XX AAV88340;

DT 14-JUL-2000 (first entry)

XX C-terminal peptide fragment of bovine alpha-S2 casein.

XX Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
KM antibacterial peptide; lactoferrin; antiviral; antitumor activity.

XX Bos sp.

PN WO200015655-A1.

XX 23-MAR-2000.

PF 15-SEP-1999; 99WO-EP07002.

PR 15-SEP-1998; 98EP-0203107.

XX 08-JUN-1999; 99EP-0201815.

XX (NIZO-) NIZO FOOD RES.

PI Vlasier S, Recto I;

DR WPI; 2000-271377/23.

XX Novel process for producing peptides with e.g. antimicrobial activity
PT from biological fluids such as milk, whey or blood comprises contacting
PT fluid with chromatographic medium to adsorb peptide domain of interest

XX Claim 14; Page 22; 41pp; English.

XX This sequence represents a C-terminal fragment of bovine alpha-S2 casein
CC protein. The peptide is an example of a peptide with antibacterial
CC activity that can be produced by the process of the invention. The
CC invention relates to a process for producing peptides from biological
CC fluids. The process comprises chromatography of the biological fluid, in
CC situ hydrolysis of selectively bound peptides, washing to remove unbound
CC peptide, and elution of the peptides of interest. The process is used for
CC producing peptides from biological fluids, such as milk, whey or blood.
CC For example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumor activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptides with a
CC purified activity of interest without the need for intermediate
CC purification of the precursor protein.

XX SQ Sequence 25 AA;

Query Match 100.0%; Score 139; DB 21; Length 25;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMWIOPKTKVIPYRYL 25
Db 1 VYQHOKAMKPMWIOPKTKVIPYRYL 25

RESULT 3
AAW32219

ID AAW32219 standard; peptide; 31 AA.

XX AAW32219;

DT 03-FEB-1998 (first entry)

XX Alpha-S2 casein precursor C-terminal peptide fragment #5.

XX Alpha-S2 casein precursor; growth promoting; mitogenic assay;

XX platelet-derived growth factor; insulin-like derived growth factor;

XX Synthetic.

PN WO9716460-A1.

PD 09-MAY-1997.

PF 31-OCT-1996; 96WO-GB02658.

PR 31-OCT-1995; 95GB-0022302.

XX (UYL-) UNIV LIVERPOOL.

PI Liu Q, Smith JA, Wilkinson MC;

DR WPI; 1997-272048/24.

XX Manufacture of medicament or foodstuff for promoting growth - using
PT peptide(s) with a sequence identical to the C-terminal end of an
PT alpha-S2 casein precursor

XX Claim 12; Page 21; 33pp; English.

XX The present sequence, which is substantially identical to the C-terminal
CC end of an alpha-S2 casein precursor, was found after storage in PBS to
CC exhibit growth promoting activity for rat mammary fibroblast cell line
CC (Rama 27), which is not significantly stimulated by platelet-derived
CC growth factor or insulin-like growth factor. The activity of the peptide
CC increased when maintained at alkaline pH. By way of contrast, alpha-S2
CC casein was inactive in a mitogenic assay. This peptide may be used in the
CC manufacture of a medicament or foodstuff for promoting growth in
CC humans or animals.

XX SQ Sequence 31 AA;

Query Match 100.0%; Score 139; DB 18; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.5e-13;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYQHOKAMKPMWIOPKTKVIPYRYL 25
Db 7 VYQHOKAMKPMWIOPKTKVIPYRYL 31

RESULT 4
AAE17466

ID AAE17466 standard; peptide; 31 AA.

XX AAE17466;

XX

DT 22-APR-2002 (first entry)
 XX Bovine alpha-S2 casein precursor protein C-terminal fragment #7.
 DE
 XX
 XX Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 XX chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Bos sp.
 XX
 PN MO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 PS WPI; 2002-154690/20.
 DR
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 XX Claim 9; Page 22; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is a peptide
 CC derived from the C-terminal of bovine alpha-S2 casein precursor protein.
 CC
 XX
 SQ Sequence 31 AA;
 XX
 QY 1 VYQHOKAMKPMWIOPTKVIPIYRYL 25
 DB 7 VYQHOKAMKPMWIOPTKVIPIYRYL 31

Query Match 100.0%; Score 139; DB 23; Length 31;
 Best Local Similarity 100.0%; Pred. No. 2.5e-13;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
 AAM32220
 ID AAM32220 standard; protein; 222 AA.
 XX
 AC AAM32220;
 XX
 DT 03-FEB-1998 (first entry)
 XX
 XX Bovine alpha-S2 casein precursor.
 DE
 XX
 XX Bovine alpha-S2 casein precursor; growth promoting; mitogenic assay;
 KM platelet-derived growth factor; insulin-like derived growth factor;
 XX
 OS Bos taurus.
 XX
 PN WO9716460-A1.
 XX
 PD 09-MAY-1997.
 XX

PF 31-OCT-1996; 96WO-GB02658.
 XX
 PR 31-OCT-1996; 95GB-0022302.
 XX
 PA (UYLI-) UNIV LIVERPOOL.
 XX
 PI Liu Q, Smith JA, Wilkinson MC;
 XX
 PS WPI; 1997-272048/24.
 DR
 XX
 PT Manufacture of medicament or foodstuff for promoting growth - using
 PT peptide(s) with a sequence identical to the C-terminal end of an
 PT alpha-S2 casein precursor
 XX
 PS Disclosure; Page 3; 33pp; English.
 XX
 CC The present sequence represents bovine alpha-S2 casein precursor.
 CC Peptides having an amino acid sequence which is substantially identical
 CC to the C-terminal end of an alpha-S2 casein precursor, are used for
 CC manufacture of a medicament or foodstuff for promoting growth in humans
 CC or animals.
 CC
 XX
 SQ Sequence 222 AA;
 XX
 QY 1 VYQHOKAMKPMWIOPTKVIPIYRYL 25
 DB 198 VYQHOKAMKPMWIOPTKVIPIYRYL 222

Query Match 100.0%; Score 139; DB 18; Length 222;
 Best Local Similarity 100.0%; Pred. No. 2.3e-12;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6
 AAE17468
 ID AAE17468 standard; Protein; 222 AA.
 XX
 AC AAE17468;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Bovine alpha-S2 casein precursor protein.
 XX
 KM Bovine; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 XX chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Bos sp.
 XX
 PN MO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 PS WPI; 2002-154690/20.
 DR
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 XX Claim 8; Page 6; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and

CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is bovine
 CC alpha-S2 casein precursor protein.

XX
 XX Sequence 222 AA;

Query Match 100.0%; Score 139; DB 23; Length 222;
 Best Local Similarity 100.0%; Pred. No. 2.3e-12;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VYOHQKAMKPMIOPTKTVIPYRYL 25
 DB 198 VYOHQKAMKPMIOPTKTVIPYRYL 222

RESULT 7

AA60481 standard; peptide; 24 AA.

AC AAR60481;
 DT 19-APR-1995 (first entry)

XX Peptide used in tranquilliser.

KW Peptide; tranquilliser; food; foodstuff; mental stabilisation.

OS Synthetic.

XX JF06211689-A.

PN 02-AUG-1994.

XX 19-JAN-1993; 93JP-0024811.

XX 19-JAN-1993; 93JP-0024811.

XX (KANE) KANEBO LTD.

DR WPI; 1994-283276/35.

XX Synthetic peptide(s) used in tranquilliser - also used in the
 PT production of a food for mental stabilisation.

XX Claim 3; Page 1; 5pp; Japanese.

CC The peptide may be used as a component of a tranquilliser which may
 CC be administered orally and is low in toxicity. The peptide may also
 CC be used as a component of a foodstuff which may be used for mental
 CC stabilisation. See also AAR60480.

XX Sequence 24 AA;

Query Match 97.1%; Score 135; DB 15; Length 24;
 Best Local Similarity 100.0%; Pred. No. 7.3e-13;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VYOHQKAMKPMIOPTKTVIPYRY 24
 DB 1 VYOHQKAMKPMIOPTKTVIPYRY 24

RESULT 8

AAV8339 standard; peptide; 24 AA.

XX
 AC AAV8339;
 XX

DT 14-JUL-2000 (first entry)
 XX Internal peptide fragment of bovine alpha-S2 caesin #3.

KW Alpha-S2 caesin; peptide production; biological fluid; milk; whey; blood;
 KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.

OS Bos sp.

XX MO200015655-A1.

PN 23-MAR-2000.

XX 15-SEP-1999; 99WO-EP07002.

XX 15-SEP-1998; 98EP-0203107.

XX 08-JUN-1999; 99EP-0201815.

XX (NIZO-) NIZO FOOD RES.

XX Visser S, Recto I;

XX WPI; 2000-271377/23.

PT Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest

XX Claim 14; Page 22; 41pp; English.

CC This sequence represents an internal fragment of bovine alpha-S2 caesin
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

XX Sequence 24 AA;

Query Match 97.1%; Score 135; DB 21; Length 24;
 Best Local Similarity 100.0%; Pred. No. 7.3e-13;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 VYOHQKAMKPMIOPTKTVIPYRY 24
 DB 1 VYOHQKAMKPMIOPTKTVIPYRY 24

RESULT 9
 AAW93885 standard; peptide; 26 AA.

XX AAW93885;

XX 25-JUN-1999 (first entry)

XX Bifidobacterium bifidus stimulating peptide 21.

XX Bifidogenic peptide; protease; treatment; microbe-related disease;
 KW bacterial fungi; yeast; protozoa; virus; mycoplasma; filaria; plasmodia;
 KW infection; inflammation; microbial induced tumour; degenerative disorder;
 KW diarrhoea; colic; oral microflora; intestinal microflora; caries;

KM vaginal microflora.
 XX Bifidobacterium bifidus.
 OS WO914231-A2.
 XX 25-MAR-1999.
 PD 16-SEP-1998; 98WO-EP05899.
 XX 11-FEB-1998; 98DE-1005385.
 PR 16-SEP-1997; 97DE-1040604.
 XX (FORS/) FORSSMANN W.
 PA Forssmann W, Liepke C, Zucht H;
 PI MPI; 1999-244022/20.
 DR
 XX
 PT Milk-derived peptides that stimulate Bifidobacterium bifidus
 PS Claim 2; Page 3; 25pp; German.
 XX
 CC This invention describes milk-derived bifidogenic peptides and their
 CC active derivatives or fragments, and combinations of them produced by
 CC chemical coupling. Such are produced from bovine or human milk by
 CC treatment for 2 hr with proteases, then centrifuging to remove fat and
 CC acidifying to pH 2 to precipitate proteins. The solution phase is then
 CC subjected to reverse-phase high performance liquid chromatography (HPLC)
 CC and cation-exchange HPLC, the fractions adjusted to salt content below
 CC 25 mM (by dialysis or reverse-phase HPLC) and tested for activity by
 CC growing Bifidobacterium bifidus and Escherichia coli in presence of the
 CC fractions. Those fractions for which (Bw-B0) - (Bw-E0) is at least 0.15
 CC are selected where Bw = germ count after 16 hr culture of B. bifidus in
 CC 50% Bliker broth containing peptide at 0.2 mg/ml, B0 = germ count under
 CC similar conditions in a peptide-free control, Bw = germ count after 16
 CC hr culture of E. coli in 3 g/l tryptic broth containing peptide at
 CC 0.2 mg/ml, E0 = germ count under similar conditions in a peptide-free
 CC control. The peptides AAW93665-W9388 are used to treat microbe-related
 CC diseases caused by bacteria, fungi, yeast, protozoa, viruses,
 CC mycoplasma, filaria and plasmodia, e.g. infections, inflammation,
 CC microbially induced tumours or degenerative disorders, diarrhoea, colic,
 CC abnormalities in oral, intestinal or vaginal microflora, or caries.
 CC
 SQ Sequence .26 AA;
 Query Match 92.4%; Score 128.5; DB 20; Length 26;
 Best Local Similarity 96.2%; Pred. No. 7.2e-12;
 Matches 25; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1 VYQHOKAMKPMWIOPTKVIPIY 25
 DB 1 VYQHOKAMKPMWIOPTKVIPIY 26
 RESULT 10
 AAY88338
 ID AAY88338 standard; peptide; 21 AA.
 AC AAY88338;
 XX
 DT 14-JUL-2000 (first entry)
 XX
 DE Internal peptide fragment of bovine alpha-S2 casein #2.
 XX
 KM Alpha-S2 casein; peptide production; biological fluid; milk; whey; blood;
 KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
 XX
 OS Bos sp.
 XX WO200015655-A1.
 PN 23-MAR-2000.
 PD

XX
 PF 15-SEP-1999; 99WO-EP07002.
 XX
 PR 15-SEP-1998; 98EP-0203107.
 PR 08-JUN-1999; 99EP-0201815.
 XX (NIZO-) NIZO FOOD RES.
 PA
 PI Visser S, Reclio I;
 XX MPI; 2000-271377/23.
 DR
 XX
 PT Novel process for producing peptides with e.g. antimicrobial activity
 PT from biological fluids such as milk, whey or blood comprises contacting
 PT fluid with chromatographic medium to adsorb peptide domain of interest
 PT
 XX
 PS Claim 14; Page 22; 41pp; English.
 XX
 CC This sequence represents an internal fragment of bovine alpha-S2 casein
 CC protein. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.
 CC
 SQ Sequence 21 AA;
 Query Match 85.6%; Score 119; DB 21; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VYQHOKAMKPMWIOPTKVIPIY 21
 DB 1 VYQHOKAMKPMWIOPTKVIPIY 21
 RESULT 11
 AAE17469
 ID AAE17469 standard; Protein; 223 AA.
 AC AAE17469;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Alpha-S2 casein precursor (alpha-S2-CN) protein #1.
 XX
 KM Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;
 KW fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;
 KW toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Unidentified.
 XX WO200202133-A2.
 PN 10-JAN-2002.
 PD 13-JUN-2001; 2001WO-GB02601.
 PF 30-JUN-2000; 2000GB-0016189.
 PR (PEPS-) PEPSYN LTD.
 PA
 XX

PI Smith JA;
 XX WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Disclosure; Page 8; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is alpha-S2
 CC casein precursor (alpha-S2-CN) protein.
 CC
 XX
 SQ Sequence 223 AA;
 Query Match 82.7%; Score 115; DB 23; Length 223;
 Best Local Similarity 84.0%; Pred. No. 7.5e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 1 VYOHQKMKPWIOPKTVIPYRYL 25
 Db 199 VDOHQKMKPWIOPKTVIPYRYL 223
 |||||
 RESULT 12
 ID AAE17470 standard; Protein; 223 AA.
 XX
 AC AAE17470;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Goat alpha-S2 casein E precursor protein.
 XX
 KM Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2

CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein E precursor protein.
 CC
 XX
 SQ Sequence 223 AA;
 Query Match 82.7%; Score 115; DB 23; Length 223;
 Best Local Similarity 84.0%; Pred. No. 7.5e-09;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 1 VYOHQKMKPWIOPKTVIPYRYL 25
 Db 199 VDOHQKMKPWIOPKTVIPYRYL 223
 |||||
 RESULT 13
 ID AAE17471 standard; Protein; 223 AA.
 XX
 AC AAE17471;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Capra hircus alpha-S2 casein C precursor protein.
 XX
 KM Goat; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
 KM collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
 KM chewing gum; cosmetic; wrinkling; periodontal disease.
 XX
 OS Capra hircus.
 XX
 WO200202133-A2.
 XX
 PD 10-JAN-2002.
 XX
 PF 13-JUN-2001; 2001WO-GB02601.
 XX
 PR 30-JUN-2000; 2000GB-0016189.
 XX
 PA (PEPS-) PEPSYN LTD.
 XX
 PI Smith JA;
 XX
 DR WPI; 2002-154690/20.
 XX
 PT Use of peptide or its derivative containing an amino acid sequence in
 PT alpha-S2 casein precursor in the manufacture of a medicament for
 PT alleviating or preventing periodontal disease and an effect of aging in
 PT skin -
 XX
 PS Claim 8; Page 8; 27pp; English.
 XX
 CC The invention relates to a composition comprising a peptide or its
 CC derivative. The peptide contains an amino acid sequence from alpha-S2
 CC casein precursor. The peptides stimulate the growth of fibroblasts, and
 CC thus the synthesis and secretion of collagen. The peptides also stimulate
 CC the growth of keratinocytes, which aid in the formation and regeneration
 CC of skin surface. The peptide is useful in the manufacture of a medicament
 CC in the form of a toothpaste or a chewing gum, for alleviating or
 CC preventing periodontal disease and a medicament in the form of a cosmetic
 CC composition for alleviating or preventing an effect of aging,
 CC particularly wrinkling of the skin. The present sequence is goat
 CC alpha-S2 casein C precursor protein.
 CC
 XX
 SQ Sequence 223 AA;
 Query Match 82.7%; Score 115; DB 23; Length 223;

Best Local Similarity 84.0%; Pred. No. 7.5e-09;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKRWIOPTKTVIPYRYL 25
DB 199 VDOHQAMKRWIOPTKTNALPYRYL 223

RESULT 14

AAE17473
ID AAE17473 standard; Protein; 223 AA.

AC AAE17473;

DT 22-APR-2002 (first entry)

DE Sheep alpha-S2 casein precursor protein.

KW Sheep; alpha-S2 casein; dermatological; antiinflammatory; fibroblast;
collagen; keratinocyte; skin regeneration; medicament; toothpaste; aging;
chewing gum; cosmetic; wrinkling; periodontal disease.

OS Ovis sp.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

XX Use of peptide or its derivative containing an amino acid sequence in

PT alpha-S2 casein precursor in the manufacture of a medicament for

PT alleviating or preventing periodontal disease and an effect of aging in

PT skin -

PS Claim 8; Page 9; 27pp; English.

XX The invention relates to a composition comprising a peptide or its

CC derivative. The peptide contains an amino acid sequence from alpha-S2

CC casein precursor. The peptides stimulate the growth of fibroblasts, and

CC thus the synthesis and secretion of collagen. The peptides also stimulate

CC the growth of keratinocytes, which aid in the formation and regeneration

CC of skin surface. The peptide is useful in the manufacture of a medicament

CC in the form of a toothpaste or a chewing gum, for alleviating or

CC preventing periodontal disease and a medicament in the form of a cosmetic

CC composition for alleviating or preventing an effect of aging,

CC particularly wrinkling of the skin. The present sequence is sheep

CC alpha-S2 casein precursor protein.

XX SQ Sequence 223 AA;

Query Match 82.7%; Score 115; DB 23; Length 223;
Best Local Similarity 84.0%; Pred. No. 7.5e-09;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKRWIOPTKTVIPYRYL 25
DB 199 VDOHQAMKRWIOPTKTNALPYRYL 223

RESULT 15

AAE17475
ID AAE17475 standard; Protein; 223 AA.

AC AAE17475;

XX 22-APR-2002 (first entry)

DT Alpha-S2 casein precursor (alpha-S2-CN) protein #2.

XX

DE Alpha-S2 casein; alpha-S2-CN; dermatological; antiinflammatory; cosmetic;

KW fibroblast; collagen; keratinocyte; skin regeneration; medicament; aging;

KW toothpaste; chewing gum; cosmetic; wrinkling; periodontal disease.

XX

OS Unidentified.

PN WO200202133-A2.

PD 10-JAN-2002.

PF 13-JUN-2001; 2001WO-GB02601.

PR 30-JUN-2000; 2000GB-0016189.

PA (PEPS-) PEPSYN LTD.

PI Smith JA;

DR WPI; 2002-154690/20.

XX

PT Use of peptide or its derivative containing an amino acid sequence in

PT alpha-S2 casein precursor in the manufacture of a medicament for

PT alleviating or preventing periodontal disease and an effect of aging in

PT skin -

PS Disclosure; Page 9; 27pp; English.

XX

CC The invention relates to a composition comprising a peptide or its

CC derivative. The peptide contains an amino acid sequence from alpha-S2

CC casein precursor. The peptides stimulate the growth of fibroblasts, and

CC thus the synthesis and secretion of collagen. The peptides also stimulate

CC the growth of keratinocytes, which aid in the formation and regeneration

CC of skin surface. The peptide is useful in the manufacture of a medicament

CC in the form of a toothpaste or a chewing gum, for alleviating or

CC preventing periodontal disease and a medicament in the form of a cosmetic

CC composition for alleviating or preventing an effect of aging,

CC particularly wrinkling of the skin. The present sequence is alpha-S2

CC casein precursor (alpha-S2-CN) protein.

XX SQ Sequence 223 AA;

Query Match 82.7%; Score 115; DB 23; Length 223;

Best Local Similarity 84.0%; Pred. No. 7.5e-09;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VYQHOKAMKRWIOPTKTVIPYRYL 25

DB 199 VDOHQAMKRWIOPTKTNALPYRYL 223

Search completed: July 30, 2003, 16:23:27
Job time : 49.256 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 49.1964 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-5

Perfect score: 164
Sequence: 1 PEWSKCYQWRMRKLGAPSIICIRRTSA 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_23:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phase:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*
16: sp_virus:*
17: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	61.0	234	4	Q8IX02
2	99	60.4	711	4	Q8ICD2
3	97	59.1	711	4	Q8IZH6
4	97	59.1	711	4	Q8IU92
5	91	55.5	33	6	Q9TR80
6	84	51.2	707	11	Q8CBA0
7	69	42.1	704	6	Q8MWN8
8	68	41.5	38	4	Q9UCY5
9	60.5	36.9	700	11	Q9DBD0
10	58	35.4	711	6	Q9XT72
11	57.5	35.1	700	11	Q8VC96
12	54.5	33.2	192	2	Q8KL39
13	54	32.9	821	2	Q8KPR9
14	53.5	32.6	736	6	Q97490
15	53	32.3	48	6	Q77558
16	53	32.3	1361	12	Q8BF56

17	53	32.3	1449	12	Q65974	Q65974 caesava com
18	52	31.7	199	11	Q8R011	Q8R011 mus musculu
19	51.5	31.4	1795	4	Q14674	Q14674 homo sapien
20	51.5	31.4	4437	2	Q8GGP3	Q8GGP3 streptomyc
21	51	31.1	489	16	P96223	P96223 mycobacteri
22	50	30.5	54	6	Q9TQV7	Q9TQV7 equus cabal
23	50	30.5	154	16	Q8PF84	Q8PF84 xanthomonas
24	50	30.5	511	11	Q8BVH0	Q8BVH0 mus musculu
25	50	30.5	522	5	Q9VWL4	Q9VWL4 drosophila
26	50	30.5	643	11	Q8BV16	Q8BV16 mus musculu
27	50	30.5	1207	4	Q8TE59	Q8TE59 homo sapien
28	49.5	30.2	1536	4	Q9C014	Q9C014 homo sapien
29	49	29.9	130	17	Q8TVE4	Q8TVE4 methanopyru
30	49	29.9	195	12	Q9IW22	Q9IW22 hydra angae r
31	49	29.9	258	5	Q9VAF6	Q9VAF6 drosophila
32	49	29.9	405	16	Q92F81	Q92F81 listeria in
33	49	29.9	408	16	Q8YAB4	Q8YAB4 listeria mo
34	49	29.9	507	16	Q9A7G8	Q9A7G8 caulobacter
35	49	29.9	558	2	Q936J3	Q936J3 rhodospirill
36	49	29.9	1721	5	Q93547	Q93547 caenorhabd
37	48.5	29.6	71	16	Q8ZJ54	Q8ZJ54 yersinia pe
38	48.5	29.6	132	16	Q8CLO9	Q8CLO9 yersinia pe
39	48.5	29.6	518	16	Q9L1H4	Q9L1H4 streptomyce
40	48	29.3	271	16	Q98JP8	Q98JP8 rhizobium i
41	48	29.3	352	7	Q9XS13	Q9XS13 aulonocara
42	48	29.3	362	16	Q8XQ97	Q8XQ97 raietonia s
43	48	29.3	393	10	Q9ZTP0	Q9ZTP0 oryza sativ
44	48	29.3	402	10	Q9ZRH8	Q9ZRH8 oryza sativ
45	48	29.3	424	16	Q98FV3	Q98FV3 rhizobium i

ALIGNMENTS

RESULT 1

ID Q8IX02 PRELIMINARY; PRT; 234 AA.
AC Q8IX02;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE Lactoferrin (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu Z., Li N.;
RT "Human lactoferrin gene, 5' regulation region and exons 1-6.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF508798; ANO14686.1; -.
FT
FT NON_TER 234
SQ SEQUENCE 234 AA; 25671 MW; 2A08DAFDB8780C7 CRC64;

Query Match 61.0%; Score 100; DB 4; Length 234;
Best Local Similarity 57.1%; Pred. No. 66-07; Mismatches 6; Indels 0; Gaps 0;

Matches 16; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 1 PEWSKCYQWRMRKLGAPSIICIRRTS 28
||:|||||:|||||:|||||:|||||:|
Db 34 PEATKCFQWRMRKRGPPVSCIKRDS 61
RESULT 2
Q8TCD2
ID Q8TCD2; PRELIMINARY; PRT; 711 AA.
AC Q8TCD2;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE Lactotransferrin.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 DR EMBL: BC022347; AAH22347.1; -
 DR InterPro: IPR001156; Transferrin.
 DR Pfam: PF00405; Transferrin; 2.
 DR SMART: SM00094; TR_FER; 2.
 DR PROSITE: PS00205; TRANSFERRIN_1; 2.
 DR PROSITE: PS00206; TRANSFERRIN_2; 2.
 DR PROSITE: PS00207; TRANSFERRIN_3; 2.
 KM Glycoprotein; Iron transport; Metal-binding; Transport.
 SQ SEQUENCE 711 AA; 78327 MW; 1B9C7BE097C45FAF CRC64;

Query Match 60.4%; Score 97; DB 4; Length 711;
 Best Local Similarity 57.1%; Pred. No. 2.5e-06;
 Matches 16; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 PEMSKCYQWQRRKXGAPSTICIRRTS 28
 ||:|||||:|||||:|||||:|||||:
 DB 34 PEATKCFQWQRRKXGAPSTICIRRTS 61

RESULT 3
 ID 081ZH6 PRELIMINARY; PRT; 711 AA.
 AC 081ZH6;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Lactoferrin.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kaplan J.B.; Fine D.H.;
 RT "Characterization of an amino acid polymorphism in the antibacterial
 domain of human lactoferrin."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY137470; AAN11304.1; -
 SQ SEQUENCE 711 AA; 78396 MW; 547AB9423C27CE67 CRC64;

Query Match 59.1%; Score 97; DB 4; Length 711;
 Best Local Similarity 53.6%; Pred. No. 5.1e-06;
 Matches 15; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 PEMSKCYQWQRRKXGAPSTICIRRTS 28
 ||:|||||:|||||:|||||:|||||:
 DB 34 PEATKCFQWQRRKXGAPSTICIRRTS 61

RESULT 4
 ID 081U92 PRELIMINARY; PRT; 711 AA.
 AC 081U92;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Lactotransferrin precursor (Lactoferrin).
 GN LTF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Shi Y.-O.; Zhang Y.; Zheng Y.-M.;
 RT "Homo sapiens Lactotransferrin Gene: cDNA cloning and Sequence
 Analysis."
 RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Semenal vesicle;
 RA Baskar Singh S.; Saravanan K.; Paramesivam M.; Srinivasan A.,
 RA Singh T.P.;
 RT "Homo sapiens lactoferrin (HLF) mRNA."
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY165046; AAN63998.1; -
 DR EMBL: AY178998; AAN75578.2; -
 KW Signal.
 FT SIGNAL.
 FT CHAIN 1 19 POTENTIAL.
 FT CHAIN 20 711 LACTOTRANSFERRIN.
 SQ SEQUENCE 711 AA; 78382 MW; 547BFC42C9267B67 CRC64;

Query Match 59.1%; Score 97; DB 4; Length 711;
 Best Local Similarity 53.6%; Pred. No. 5.1e-06;
 Matches 15; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 1 PEMSKCYQWQRRKXGAPSTICIRRTS 28
 ||:|||||:|||||:|||||:|||||:
 DB 34 PEATKCFQWQRRKXGAPSTICIRRTS 61

RESULT 5
 ID 09TR80 PRELIMINARY; PRT; 33 AA.
 AC 09TR80;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE Lactoferrin (Fragment).
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 NCBI_TaxID=9940;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=95127729; PubMed=7827104;
 RA Qian Z.Y.; Jolles P.; Migliore-Samour D.; Fiat A.M.;
 RL Biochim. Biophys. Acta 1243:25-32(1995).
 DR HSP: O77698; ICE2.
 DR InterPro: IPR001156; Transferrin.
 DR Pfam: PF00405; Transferrin; 1.
 SQ SEQUENCE 33 AA; 3914 MW; D1904CAB15A73961 CRC64;

Query Match 55.5%; Score 91; DB 6; Length 33;
 Best Local Similarity 78.9%; Pred. No. 1.9e-06;
 Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PEMSKCYQWQRRKXGAP 19
 ||:|||||:|||||:|||||:|||||:
 DB 14 PEGSKCYQWQRRKXGAP 32

RESULT 6
 ID 08CBA0 PRELIMINARY; PRT; 707 AA.
 AC 08CBA0;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Lactotransferrin.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Bone;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The PANTOM Consortium
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs";
 RL Nature 420:563-573(2002).
 DR EMBL; AK036491; BAC29450.1;
 SQ SEQUENCE 707 AA; 77837 MW; E1B32F5FD8748A0F CRC64;
 Query Match 51.2%; Score 84; DB 11; Length 707;
 Best Local Similarity 44.4%; Pred. No. 0.00044;
 Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;
 QY 2 EMSKCYQWQRMKLGAPSTICRTS 28
 DB 33 EEEKCIHQWEMKXVGPPLSYKXS 59
 RESULT 7
 O8MMN8 PRELIMINARY; PRT; 704 AA.
 ID O8MMN8;
 AC O8MMN8;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Lactoferrin (Fragment).
 GN LTF.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxId=9823;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Wang S.-R., Lin T.-Y., Weng C.-N.;
 RT "Isolation and expression of porcine milk lactoferrin";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBD databases.
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC EMBL; L77887; AAL40161.1;
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; Transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.
 KW Glycoprotein; Iron transport; Metal-binding; Transport.
 FT NON TER 1
 SQ SEQUENCE 704 AA; 77681 MW; 64EE769F7503CC32 CRC64;
 Query Match 42.1%; Score 69; DB 6; Length 704;
 Best Local Similarity 51.9%; Pred. No. 0.077;
 Matches 14; Conservative 4; Mismatches 7; Indels 2; Gaps 1;
 QY 2 EMSKCYQWQRMKLGAPSTICRTS 28
 DB 34 EYKCKRQWQSKLRIN-PIFCTRAS 58
 RESULT 8
 O9UCY5 PRELIMINARY; PRT; 38 AA.
 ID O9UCY5;
 AC O9UCY5;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Lactoferrin homolog (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96081613; PubMed=8551695;
 RA Sato I.;
 RT "Characterization of the 84-kDa protein with ABH activity in human
 RT seminal plasma";
 RT Upn. J. Legal Med. 49:281-293(1995).
 DR HSSP; 077811; 1B1X.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; Transferrin; 1.
 SQ SEQUENCE 38 AA; 4459 MW; 0402F490B5EBDB5 CRC64;
 Query Match 41.5%; Score 68; DB 4; Length 38;
 Best Local Similarity 50.0%; Pred. No. 0.006;
 Matches 12; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
 QY 1 EWSKCYQWQRMKLGAPSTIC 24
 DB 15 PEADKXFWQRMKRGVXSXI 38
 RESULT 9
 O9DBD0 PRELIMINARY; PRT; 700 AA.
 ID O9DBD0;
 AC O9DBD0;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE 130001702Rik protein.
 GN 130001702Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Liver;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt R., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,
 RA Blake J., Boftelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.D., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guslinich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima U., Mazzarelli J., Momberts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 DR EMBL; AK005035; BAB23762.1; -.
 DR HSSP; P19134; 1TPD.
 DR MGD; MGI:1919025; 130001702Rik.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; Transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 1.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.

SQ SEQUENCE 700 AA; 76765 MW; 38C991D1021AE548 CRC64;
 Query Match 36.2%; Score 60.5; DB 11; Length 700;
 Best Local Similarity 36.7%; Pred. No. 1.4;
 Matches 11; Conservative 9; Mismatches 7; Indels 3; Gaps 1;

QY 2 EWSKCYOMORRMKRL--GAPSTICIRRTS 28
 34 EATKCSSFRDMKKVLPAGGPVAVTCVRKMS 63

RESULT 10
 Q9XT72 PRELIMINARY; PRT; 711 AA.
 ID Q9XT72
 AC Q9XT72
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Serum iron transport protein transferrin.
 GN TRFE.
 OS Trichosurus vulpecula (Brush-tailed possum).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
 OC NCBI_TaxID=9337;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=99227120; PubMed=10209259;
 RA Demmer J., Staetuk S.J., Adamaki F.M., Grigor M.R.;
 RT "Cloning and expression of the transferrin and ferritin genes in a
 marsupial, the brushtail possum (Trichosurus vulpecula).";
 RL Biochim. Biophys. Acta 1445:65-74(1999).
 CC -1- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 DR EMBL: AF092510; AAD3831.1; -;
 DR HSSP: P56410; 1A0V.
 DR InterPro: IPR001156; Transferrin.
 DR Pfam: PF00405; transferrin; 2.
 DR PRINTS: PRO0422; TRANSFERRIN.
 DR SMART: SM00094; TR_FER; 2.
 DR PROSITE: PS00206; TRANSFERRIN_2; 1.
 DR PROSITE: PS00207; TRANSFERRIN_3; 1.
 KM Glycoprotein; Iron transport; Metal-binding; Transport.
 SQ SEQUENCE 711 AA; 77764 MW; 35CA4994DD95631F CRC64;

Query Match 35.4%; Score 58; DB 6; Length 711;
 Best Local Similarity 33.3%; Pred. No. 3.4;
 Matches 9; Conservative 7; Mismatches 11; Indels 0; Gaps 0;

QY 2 EWSKCYOMORRMKRLGAPSTICIRRTS 28
 33 ELKKCNFEKAMNSINHGLACTIKTS 59

RESULT 11
 Q8VC96 PRELIMINARY; PRT; 700 AA.
 ID Q8VC96
 AC Q8VC96
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE RIKEN cDNA 1300017J02 gene.
 GN 1300017J02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Strausberg R.;

RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC021390; AAH21390.1; -;
 DR MGD: MGI:1919025; 1300017J02RIK.
 DR InterPro: IPR001156; Transferrin.
 DR Pfam: PF00405; transferrin; 2.
 DR PRINTS: PRO0422; TRANSFERRIN.
 DR SMART: SM00094; TR_FER; 2.
 DR PROSITE: PS00205; TRANSFERRIN_1; 1.
 DR PROSITE: PS00207; TRANSFERRIN_3; 1.
 SQ SEQUENCE 700 AA; 76795 MW; 7DA4A580CB6BF37 CRC64;

Query Match 35.1%; Score 57.5; DB 11; Length 700;
 Best Local Similarity 37.0%; Pred. No. 4;
 Matches 10; Conservative 8; Mismatches 6; Indels 3; Gaps 1;

QY 5 KCYOMORRMKRL--GAPSTICIRRTS 28
 37 KCSFRDMKKVLPAGGPVAVTCVRKMS 63

RESULT 12
 Q8KL39 PRELIMINARY; PRT; 192 AA.
 ID Q8KL39
 AC Q8KL39
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein.
 GN YP044.
 OS Rhizobium etli.
 OG Plasmid symbiotic plasmid p42d.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium.
 OC NCBI_TaxID=29449;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CFN42;
 RX MEDLINE=91193195; PubMed=2013564;
 RA Girard M.L., Flores M., Brom S., Romero D., Palacios R., Davila G.;
 RT "Structural complexity of the symbiotic plasmid of Rhizobium
 leguminosarum bv. phaseoli";
 RL J. Bacteriol. 173:2411-2419(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CFN42;
 RX MEDLINE=97419521; PubMed=9274036;
 RA Ramirez-Romero M.A., Bustos P., Girard L., Rodriguez O.,
 RA Cevallos M.A., Davila G.;
 RT "Sequence, localization and characteristics of the replicator region
 of the symbiotic plasmid of Rhizobium etli";
 RL Microbiology 143:2825-2831(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CFN42;
 RA Quintero V., Cevallos M.A., Davila G.;
 RT "A site-specific recombinase and RecA are required to exert
 incompatibility towards the symbiotic plasmid of Rhizobium etli";
 RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U80928; AAMS4928.1; -;
 KM Hypothetical protein; Plasmid.
 SQ SEQUENCE 192 AA; 21513 MW; 7AC7624DFOEB32E CRC64;

Query Match 33.2%; Score 54.5; DB 2; Length 192;
 Best Local Similarity 60.0%; Pred. No. 3.1;
 Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 9 WQRRMKLGAPSTICIRRTS 28
 148 WQRRMGSGAP-VTAIRRS 166

RESULT 13
 Q8KPR9

ID 08KPS9 PRELIMINARY; PRT; 821 AA.
 AC 08KPS9;
 DT 01-OCT-2002 (TEMBLrel. 22, Created)
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
 DE Hypothetical protein.
 GN SE00025.
 OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
 NC NCB1_TaxID=1140;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC 7942;
 RA Holtman C.K., Sandoval P., Chen Y., Socias T., Mohler B.J.,
 McMurtry S., Gonzalez A., Salinas I., Golden S.S., Youderian P.;
 RT "Synecococcus elongatus PCC7942 cosmid 763";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF120853; AAC82700.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 821 AA; 91706 MW; D5221AD0ED0DBCC CRC64;

Query Match 32.3%; Score 54; DB 2; Length 821;
 Best Local Similarity 47.8%; Pred. No. 16;
 Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
 QY 1 PEWSKCYQWRMRKLGAPSIITC 23
 ID 097490 PRELIMINARY; PRT; 736 AA.
 AC 097490;
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-MAY-2003 (TEMBLrel. 23, Last annotation update)
 DE P97.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 NC NCB1_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98451505; PubMed=9780225;
 RA Kawamoto T., Pan H., Yan W., Ishida H., Usui E., Oda R., Nakamasu K.,
 Noshiro M., Kawashima-Onya Y., Fujii M., Shintani H., Okada Y.,
 Kato Y.;
 RT "Expression of membrane-bound transferrin-like protein p97 on the cell
 surface of chondrocyte";
 RL Eur. J. Biochem. 256:503-509(1998).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC EMBL; AB010995; BAA33956.1; -.
 CC HSSP; P19134; ITPD.
 DR InterPro; IPR001064; Cysteallin.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.
 DR PROSITE; PS00205; TRANSFERRIN_2; 1.
 DR PROSITE; PS00206; TRANSFERRIN_1; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.
 DR GlycoProtein; Itron transport; Metal-binding; Transport.
 KW Glycoprotein; Itron transport; Metal-binding; Transport.
 SQ SEQUENCE 736 AA; 80169 MW; F389D9F8A6AC90FC CRC64;

Query Match 32.6%; Score 53.5; DB 6; Length 736;
 Best Local Similarity 36.7%; Pred. No. 17;
 Matches 11; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

QY 1 PEWSKCYQWRMRKLG-APSIITCIRTS 29
 ID 077558 PRELIMINARY; PRT; 48 AA.
 AC 077558;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE Lactoferrin (Fragment).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 NC NCB1_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98332734; PubMed=9666128;
 RA Wang S.R., Lin J., Cheng I.C., Lin T.Y.;
 RT "Characterization and functional analysis of the porcine lactoferrin
 gene promoter";
 RL Gene 215:203-212(1998).
 DR EMBL; AF044256; AAC34369.1; -.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 1.
 FT NON TER 48
 SQ SEQUENCE 48 AA; 5569 MW; 28403BE7DE14AD78 CRC64;

Query Match 32.3%; Score 53; DB 6; Length 48;
 Best Local Similarity 57.1%; Pred. No. 1.3;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 QY 2 EWSKCYQWRMRK 15
 ID 077558 PRELIMINARY; PRT; 48 AA.
 AC 077558;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE Lactoferrin (Fragment).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 NC NCB1_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98332734; PubMed=9666128;
 RA Wang S.R., Lin J., Cheng I.C., Lin T.Y.;
 RT "Characterization and functional analysis of the porcine lactoferrin
 gene promoter";
 RL Gene 215:203-212(1998).
 DR EMBL; AF044256; AAC34369.1; -.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 1.
 FT NON TER 48
 SQ SEQUENCE 48 AA; 5569 MW; 28403BE7DE14AD78 CRC64;

Search completed: July 30, 2003, 16:29:25
 Job time : 51.1964 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 ; Search time 9.49405 Seconds
(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-5

Perfect score: 164
Sequence: 1 PMSKCYQWQRMRKLGAPSTICIRRTSA 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	144	100.0	708	1 TRFL_CAPHI	Q29477 capra hircu
2	124	75.6	708	1 TRFL_BUBBU	O77698 bubalus bub
3	122	74.4	708	1 TRFL_BOVIN	P24627 bos taurus
4	104	63.4	708	1 TRFL_CAMDR	O94000 camelus dro
5	100	61.0	711	1 TRFL_HUMAN	P02788 homo sapien
6	84	51.2	707	1 TRFL_MOUSE	P08071 mus musculu
7	83	50.6	695	1 TRFL_HORSE	O77811 equus cabal
8	66	40.2	704	1 TRFL_PIG	P14632 sus scrofa
9	58.5	35.7	704	1 ICA_PIG	Q29545 sus scrofa
10	58	35.4	271	1 MGPB_ECO57	O8xb99 escherichia
11	58	35.4	271	1 MGPB_ECOLI	P76329 escherichia
12	58	35.4	271	1 MGPB_SALTY	O8xg91 salmonella
13	57.5	35.1	738	1 TRFL_HUMAN	P08582 homo sapien
14	52	31.7	198	1 PEBT_MOUSE	O61907 mus musculu
15	52	31.7	271	1 MGPB_ECOLI	P59286 escherichia
16	51.5	31.4	1795	1 ESP1_HUMAN	O14674 homo sapien
17	51	31.1	198	1 PEBT_RAT	O08388 rattus norv
18	50	30.5	642	1 Z398_HUMAN	O8td17 homo sapien
19	50	30.5	706	1 TRFL_HORSE	F27425 equus cabal
20	50	30.5	1207	1 AT19_HUMAN	O8te59 homo sapien
21	48	29.3	430	1 SYH_CHLPN	O9z7p1 chlamydia p
22	47.5	29.0	450	1 FEM2_RAT	O9wv77 rattus norv
23	47	28.7	198	1 PEBT_HUMAN	O9ubn1 homo sapien
24	47	28.7	704	1 TRFL_HUMAN	O12931 homo sapien
25	47	28.7	704	1 IP3T_HUMAN	O14573 homo sapien
26	46.5	28.4	695	1 TRFL_MOUSE	O9w134 oryctolagus
27	46	28.0	63	1 BD03_MOUSE	O9w134 mus musculu
28	46	28.0	413	1 YEIT_SALTY	O8z546 salmonella
29	46	28.0	413	1 YEIT_SALTY	O8z546 salmonella
30	46	28.0	488	1 VE2_HPV49	P36795 human papil
31	46	28.0	647	1 TOP5_VIBCH	O9kqf5 vibrio chol
32	46	28.0	1082	1 A3B2_HUMAN	O13367 homo sapien
33	46	28.0	1210	1 AT19_MOUSE	P59509 mus musculu

34	45.5	27.7	641	1 SCAB_RABIT	O97742 oryctolagus
35	45.5	27.7	698	1 TRFL_HUMAN	P02787 homo sapien
36	45.5	27.7	967	1 AT51_HUMAN	O9ub18 homo sapien
37	45	27.4	199	1 NMHA_PSECU	P27764 pseudomonas
38	45	27.4	310	1 TF2B_METTH	O26571 methanobact
39	45	27.4	368	1 HA14_MOUSE	P14427 mus musculu
40	45	27.4	1385	1 RRPD_PIAMV	O07518 plantago as
41	45	27.4	1490	1 CRK7_HUMAN	O9uyv4 homo sapien
42	44.5	27.1	937	1 SYL_METTH	O27552 methanobact
43	44.5	27.1	1793	1 YCPL_LOTUA	O9dbn6 lotus japon
44	44	26.8	142	1 RRPD_ORNBR	P31843 oenothera b
45	44	26.8	236	1 YIHL_ECOLI	P32133 escherichia

ALIGNMENTS

RESULT 1
ID TRFL_CAPHI STANDARD; PRT; 708 AA.
AC Q29477; Q29479;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactotransferrin precursor (Lactoferrin).
LN LTP.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP TISSUE=Mammary gland;
RC TISSUE=Mammary gland;
RA Lee T., Yu S., Kim S., Lee K., Yu D.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP TISSUE=Mammary gland;
RC MEDLINE=9438047; PubMed=8093048;
RA le Provost F., Nocard M., Guerin G., Martin P.;
RT "Characterization of the goat lactoferrin cDNA. Assignment of the
relevant locus to bovine U12 syntenic group."
RL Biochem. Biophys. Res. Commun. 203:1324-1332(1994).
CC -!- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
OF AN ANION, USUALLY BICARBONATE.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: US3857; AAA97958.1; -
CC EMBL: X78902; CAA55517.1; -
CC HSSP: O77698; 1CE2.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SMO0094; TR_PFR; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
KW signal.
FT SIGNAL 1 19 BY SIMILARITY.

```

FT CHAIN 20 708 LACTOTRANSFERRIN.
FT REPEAT 20 363 1.
FT 364 708 2.
FT DISULFID 28 64 BY SIMILARITY.
FT DISULFID 38 55 BY SIMILARITY.
FT DISULFID 134 217 BY SIMILARITY.
FT DISULFID 176 192 BY SIMILARITY.
FT DISULFID 189 200 BY SIMILARITY.
FT DISULFID 250 264 BY SIMILARITY.
FT DISULFID 367 399 BY SIMILARITY.
FT DISULFID 377 390 BY SIMILARITY.
FT DISULFID 424 703 BY SIMILARITY.
FT DISULFID 444 666 BY SIMILARITY.
FT DISULFID 476 551 BY SIMILARITY.
FT DISULFID 500 694 BY SIMILARITY.
FT DISULFID 510 524 BY SIMILARITY.
FT DISULFID 521 534 BY SIMILARITY.
FT DISULFID 592 606 BY SIMILARITY.
FT DISULFID 644 649 BY SIMILARITY.
FT METAL 79 79 ION 1 (BY SIMILARITY).
FT METAL 111 111 ION 1 (BY SIMILARITY).
FT METAL 211 211 ION 1 (BY SIMILARITY).
FT METAL 272 272 ION 1 (BY SIMILARITY).
FT METAL 414 414 ION 2 (BY SIMILARITY).
FT METAL 452 452 ION 2 (BY SIMILARITY).
FT METAL 545 545 ION 2 (BY SIMILARITY).
FT METAL 614 614 ION 2 (BY SIMILARITY).
FT BINDING 140 140 ANION (BY SIMILARITY).
FT BINDING 482 482 ANION (BY SIMILARITY).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 56 56 I -> V (IN REF. 2).
FT CONFLICT 88 88 L -> R (IN REF. 2).
FT CONFLICT 124 124 Q -> K (IN REF. 2).
FT CONFLICT 154 154 S -> P (IN REF. 2).
FT CONFLICT 304 304 S -> R (IN REF. 2).
FT CONFLICT 414 414 D -> G (IN REF. 2).
SO SEQUENCE 708 AA; 77358 MW; F2EDA3C63539960D CRC64;

```

Query Match 100.0%; Score 164; DB 1; Length 708;
 Best Local Similarity 100.0%; Pred. No. 1.4e-16;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 PEMSKCYOMORRRKLGAPSTICIRRTSA 29
DB 33 PEMSKCYOMORRRKLGAPSTICIRRTSA 61

```

```

RESULT 2
TRFL_BUBBU STANDARD; PRT; 708 AA.
ID TRFL_BUBBU
AC 077658;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lactotransferrin precursor (lactoferrin).
GN LTF.
OS Bubalus bubalis (domestic water buffalo).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bubalus.
OX NCBI_Taxid=89462;
RN [1]
RP SEQUENCE FROM N.A.
RA Paramasivam M., Thattaiyath B.D., Kumar A., Srinivasan A.,
RT Singh T.P.;
RL "cDNA sequence of Buffalo lactoferrin."
RN [2] Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

```

```

RX MEDLINE=20003130; PubMed=10531476;
RA Karthikeyan S., Paramasivam M., Yadav S., Srinivasan A., Singh T.P.;
RT "Structure of buffalo lactoferrin at 2.5-A resolution using crystals
RT grown at 303 K shows different orientations of the N and C lobes.";
RL Acta Crystallogr. D 55:1805-1813(1999).
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ005203; CA06441.1; -.
DR PDB; 1CE2; 19-MAR-99.
DR PDB; 1BTY; 13-JAN-99.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
DR Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
DR Signal; 3d-structure.
FT CHAIN 1 19 LACTOTRANSFERRIN.
FT DISULFID 20 708
FT DISULFID 38 64
FT DISULFID 134 55
FT DISULFID 176 217
FT DISULFID 192 202
FT DISULFID 250 264
FT DISULFID 367 399
FT DISULFID 377 390
FT DISULFID 424 703
FT DISULFID 444 666
FT DISULFID 476 551
FT DISULFID 500 694
FT DISULFID 510 524
FT DISULFID 521 534
FT DISULFID 592 606
FT DISULFID 644 649
FT METAL 79 79 ION 1.
FT METAL 111 111 ION 1.
FT METAL 211 211 ION 1.
FT METAL 272 272 ION 1.
FT METAL 414 414 ION 2.
FT METAL 452 452 ION 2.
FT METAL 545 545 ION 2.
FT METAL 614 614 ION 2.
FT BINDING 140 140 ANION (POTENTIAL).
FT BINDING 482 482 ANION (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 21 22
FT STRAND 25 29
FT TURN 32 46
FT TURN 47 48
FT STRAND 53 57
FT HELIX 61 69
FT TURN 70 71

```



```

FT STRAND 75 78
FT HELIX 80 87
FT TURN 89 91
FT STRAND 93 102
FT STRAND 107 108
FT STRAND 110 118
FT TURN 119 120
FT HELIX 125 127
FT TURN 129 130
FT STRAND 132 135
FT TURN 138 139
FT TURN 141 144
FT HELIX 145 150
FT TURN 151 151
FT HELIX 152 155
FT TURN 159 161
FT HELIX 164 169
FT TURN 170 171
FT STRAND 174 176
FT TURN 178 179
FT TURN 182 184
FT HELIX 186 189
FT TURN 190 191
FT HELIX 196 198
FT TURN 199 200
FT TURN 204 205
FT TURN 207 208
FT HELIX 211 219
FT TURN 220 221
FT STRAND 225 229
FT TURN 230 231
FT HELIX 232 236
FT HELIX 240 243
FT TURN 244 245
FT STRAND 246 249
FT TURN 251 252
FT STRAND 255 257
FT HELIX 258 260
FT TURN 261 263
FT STRAND 267 270
FT STRAND 273 277
FT HELIX 283 297
FT TURN 299 300
FT TURN 307 308
FT TURN 312 313
FT TURN 321 322
FT STRAND 325 328
FT TURN 331 332
FT HELIX 335 351
FT HELIX 354 362
FT STRAND 364 369
FT HELIX 371 384
FT TURN 385 388
FT STRAND 388 393
FT HELIX 396 404
FT TURN 405 406
FT STRAND 410 413
FT HELIX 415 422
FT TURN 423 425
FT STRAND 426 434
FT TURN 438 441
FT TURN 444 446
FT STRAND 452 459
FT TURN 460 461
FT TURN 463 464
FT HELIX 467 469
FT TURN 471 472
FT STRAND 474 477
FT TURN 480 481
FT TURN 483 486
FT HELIX 487 497
FT TURN 503 505
FT STRAND 508 510

```

```

FT TURN 512 513
FT TURN 516 517
FT TURN 519 520
FT TURN 522 523
FT TURN 528 529
FT TURN 533 534
FT TURN 538 539
FT TURN 541 542
FT HELIX 544 553
FT TURN 554 555
FT STRAND 559 563
FT HELIX 564 569
FT TURN 570 570
FT HELIX 578 581
FT TURN 582 582
FT TURN 585 586
FT STRAND 588 591
FT TURN 593 594
FT STRAND 597 599
FT TURN 600 601
FT TURN 603 605
FT STRAND 609 611
FT STRAND 615 619
FT HELIX 620 637
FT TURN 639 640
FT TURN 642 647
FT TURN 650 651
FT TURN 654 655
FT TURN 662 663
FT STRAND 664 669
FT HELIX 676 680
FT HELIX 682 692
FT TURN 693 694
FT HELIX 698 705

```

Query Match Best Local Similarity 75.6%; Score 124; DB 1; Length 708;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 PEMSKCYOWMRBKLGAPSIICIR 26
Db 33 PEMKCHRWOMRKLGAPSIICVRR 58

```

```

RESULT 3
TRFL_BOVIN STANDARD; PRT; 708 AA.
AC P24627; Q29629; Q9WZY3;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactotransferrin precursor (Lactoferrin) [Contains: Lactoferricin B
(LFCIN B)].
GN LTF.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxId=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Submaxillary gland;
RX MEDLINE=91160550; Pubmed=2001696;
RA Pierce A., Colavizza D., Benaissa M., Maes P., Tartar A.,
RA Montreuil J., Spik G.,
RT "Molecular cloning and sequence analysis of bovine lactotransferrin.";
RL Eur. J. Biochem. 196;177-184(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92028986; Pubmed=1718281;
RA Goodman R.E., Schanbacher F.L.;
RT "Bovine lactoferrin mRNA: sequence, analysis, and expression in the
mammary gland.";

```



```

FT HELIX 80 87
FT TURN 89 91
FT STRAND 93 102
FT STRAND 107 108
FT STRAND 110 118
FT TURN 126 127
FT TURN 129 130
FT STRAND 133 135
FT TURN 138 139
FT TURN 141 144
FT HELIX 145 155
FT TURN 156 156
FT TURN 159 161
FT HELIX 164 169
FT TURN 170 171
FT STRAND 175 176
FT TURN 178 179
FT TURN 182 184
FT HELIX 186 188
FT TURN 190 191
FT TURN 196 197
FT TURN 199 200
FT TURN 204 205
FT TURN 207 208
FT HELIX 210 219
FT TURN 220 221
FT STRAND 225 229

```

Query Match 74.4%; Score 122; DB 1; Length 708;
 Best Local Similarity 72.4%; Pred. No. 2.3e-10;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 PEMSKCYOMORRMKKGAPSTICIRRTSA 29
 Db 33 PEMFKCRPMORRMKKGAPSTICVRAPA 61

RESULT 4
 ID TRFL CAMDR STANDARD; PRT; 708 AA.

AC 09TOM0; Q9WZS5; 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DB Lactoferrin precursor (Lactoferrin).
 GN LTF.
 OS Camelus dromedarius (Dromedary) (Arabian camel).
 OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 OX NCBI_TaxID=9838;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Small; TISSUE=Lactating mammary gland;
 RA Kappeler S.R., Ackermann M., Farah Z., Puhar Z.;
 RT "Sequence analysis of camel (Camelus dromedarius) Lactoferrin";
 RL Int. Dairy J. 9:481-486(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Paramasivam M., Srinivasan A., Singh R., Sahani M.S., Singh T.P.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC -1- SUBUNIT: Monomer (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AJ131674; CAB53387.1; -;
 CC EMBL: AF165879; AAF82241.1; -;
 CC PDB: 1DT2; 20-JUN-01.
 DR InterPro; IPR01156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FBR; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 KM Transprot; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KM Signal; 3d-structure.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 708 LACTOTRANSFERRIN.
 FT REPEAT 20 363 1.
 FT REPEAT 364 708 2.
 FT DISULFID 28 64 BY SIMILARITY.
 FT DISULFID 38 55 BY SIMILARITY.
 FT DISULFID 134 217 BY SIMILARITY.
 FT DISULFID 176 192 BY SIMILARITY.
 FT DISULFID 189 200 BY SIMILARITY.
 FT DISULFID 250 264 BY SIMILARITY.
 FT DISULFID 367 399 BY SIMILARITY.
 FT DISULFID 377 380 BY SIMILARITY.
 FT DISULFID 424 703 BY SIMILARITY.
 FT DISULFID 444 666 BY SIMILARITY.
 FT DISULFID 476 551 BY SIMILARITY.
 FT DISULFID 500 694 BY SIMILARITY.
 FT DISULFID 510 524 BY SIMILARITY.
 FT DISULFID 521 534 BY SIMILARITY.
 FT DISULFID 532 606 BY SIMILARITY.
 FT DISULFID 644 649 BY SIMILARITY.
 FT METAL 79 79 IRON 1 (BY SIMILARITY).
 FT METAL 111 111 IRON 1 (BY SIMILARITY).
 FT METAL 211 211 IRON 1 (BY SIMILARITY).
 FT METAL 272 272 IRON 1 (BY SIMILARITY).
 FT METAL 414 414 IRON 2 (BY SIMILARITY).
 FT METAL 452 452 IRON 2 (BY SIMILARITY).
 FT METAL 545 545 IRON 2 (BY SIMILARITY).
 FT METAL 614 614 IRON 2 (BY SIMILARITY).
 FT BINDING 140 140 ANION (BY SIMILARITY).
 FT BINDING 482 482 ANION (BY SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 385 385 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 537 537 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 594 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 261 261 F -> S (IN REF. 2).
 FT CONFLICT 304 304 G -> A (IN REF. 2).
 FT CONFLICT 330 330 S -> P (IN REF. 2).
 FT CONFLICT 330 330 LLS -> PLF (IN REF. 2).
 FT CONFLICT 492 494 L -> F (IN REF. 2).
 FT CONFLICT 506 506 A -> P (IN REF. 2).
 FT CONFLICT 609 609 R -> O (IN REF. 2).
 FT CONFLICT 642 642 R -> O (IN REF. 2).
 SQ SEQUENCE 708 AA; 77211 MW; 0B0C175A0B69D430 CRC64;

Query Match 63.4%; Score 104; DB 1; Length 708;
 Best Local Similarity 63.0%; Pred. No. 1.1e-07;
 Matches 17; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 2 EMSKCYOMORRMKKGAPSTICIRRTS 28
 Db 34 ESSKCAOWORRMKKGAPSTICVTKRTS 60

RESULT 5
 ID TRFL HUMAN STANDARD; PRT; 711 AA.
 AC P02788; O00756; Q16780; Q16785; Q16786; Q16789; Q96K24; Q96K25;
 AC Q9H123;

DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lactoferrin precursor (lactoferrin) [Contains: Lactoferrin A;
 DE Lactoferrin B; Lactoferrin C].
 GN LTF OR LF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=90384839; PubMed=2402455;
 RA Ray M.W., Woloshuk S.L., de Boer H.A., Pieper F.R.;
 RT "Complete nucleotide sequence of human mammary gland lactoferrin.";
 RL Nucleic Acids Res. 18:5288-5288(1990).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Cho Y.Y.;
 RL Thesis (1994), Genetic Engineering Research Institute / Taejeon, Korea.
 RN [3]
 RN SEQUENCE FROM N.A.
 RA Connolly O.M.;
 RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Liang Q., Jimenez-Flores R., Richardson T.;
 RT "Molecular cloning and sequence analysis of human lactoferrin.";
 RL Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RA Wei X., Han J., Rado T.A.;
 RT "Human neutrophil lactoferrin coding and 5' flanking region DNA
 RT sequences.";
 RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Cheng H., Chen X., Huan L.;
 RT "cDNA cloning and sequence analysis of human lactoferrin.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RX MEDLINE=2338257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueidi T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunterane P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fehey J., Helton J., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [8]
 RN SEQUENCE OF 3-711 FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=90326549; PubMed=2374734;
 RA Powell M.J., Ogden J.E.;
 RT "Nucleotide sequence of human lactoferrin cDNA.";
 RL Nucleic Acids Res. 18:4013-4013(1990).
 RN [9]
 RN SEQUENCE OF 20-711.
 RX MEDLINE=8507667; PubMed=6510420;
 RA Metz-Boutigue M.-H., Jolles J., Mazurier J., Schoentgen F.,
 RA Legrand D., Spik G., Montreuil J., Jolles P.;
 RT "Human lactoferrin: amino acid sequence and structural
 RT comparisons with other transferrins.";
 RL Eur. J. Biochem. 145:659-666(1984).
 RN [10]
 RN PRELIMINARY SEQUENCE OF 20-73; 134-171; 257-278 AND 347-530.
 RX MEDLINE=82046817; PubMed=6794640;
 RA Metz-Boutigue M.-H., Mazurier J., Jolles J., Spik G., Montreuil J.,
 RA Jolles P.;
 RT "The present state of the human lactoferrin sequence. Study and
 RT alignment of the cyanogen bromide fragments and characterization of
 RT N- and C-terminal domains.";
 RL Biochim. Biophys. Acta 670:243-254(1981).
 RN [11]
 RN SEQUENCE OF 609-711.
 RX MEDLINE=82262043; PubMed=7049727;
 RA Metz-Boutigue M.-H., Jolles J., Mazurier J., Spik G., Montreuil J.,
 RA Jolles P.;
 RT "An 88 amino acid long C-terminal sequence of human
 RT lactoferrin.";
 RL FEBS Lett. 142:107-110(1982).
 RN [12]
 RN SEQUENCE OF 436-711 FROM N.A.
 RX MEDLINE=88001031; PubMed=3477300;
 RA Rado T.A., Wei X., Benz E.J., Jr.;
 RT "Isolation of lactoferrin cDNA from a human myeloid library and
 RT expression of mRNA during normal and leukemic myelopoiesis.";
 RL Blood 70:989-993(1987).
 RN [13]
 RN SEQUENCE OF 237-711 FROM N.A.
 RA McCombie W.R., Wilson R., Chen E., Gibbs R., Zuo L., Johnson D.,
 RA Nhan M., Parnell L., Dedhia N., Ansari A., Mardis E., Schutz K.,
 RA Gao L., La Bastide M., Kaplan J., Greco T., Touchman J., Muzny D.,
 RA Chen C.N., Evans C., Fitzgerald M., See L.H., Tang M., Porcel B.M.,
 RA Dragan Y., Giacalone J., Pae A., Powell E., Solinsky K.A., Desilva U.,
 RA Diaz-Perez S., Zhou X., Yu Y., Watanabe M., Doggett N., Garcia D.,
 RA Segripari J.L.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [14]
 RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND REVISIONS.
 RX MEDLINE=90064528; PubMed=2385506;
 RA Anderson B.F., Baker H.M., Norris G.E., Rice D.W., Baker E.N.;
 RT "Structure of human lactoferrin: crystallographic structure analysis
 RT and refinement at 2.8-A resolution.";
 RL J. Mol. Biol. 209:711-734(1989).
 RN [15]
 RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RA Haridas M., Anderson B.F., Baker E.N.;
 RT "Structure of human different lactoferrin refined at 2.2-A
 RT resolution.";
 RL Acta Crystallogr. D 51:629-646(1995).
 RN [16]
 RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 20-353.
 RX MEDLINE=97156796; PubMed=9003186;
 RA Nicholson H., Anderson B.F., Bland T., Shewry S.C., Tweedie J.W.,
 RA Baker E.N.;
 RT "Mutagenesis of the histidine ligand in human lactoferrin: iron
 RT binding properties and crystal structure of the histidine-
 RT 253->methionine mutant.";
 RL Biochemistry 36:341-346(1997).
 RN [17]
 RN X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=99190892; PubMed=10089347;
 RA Sun X.L., Baker H.M., Shewry S.C., Jameson G.B., Baker E.N.;
 RT "Structure of recombinant human lactoferrin expressed in Aspergillus
 RT awamori.";

KM Signal. 1 19 BY SIMILARITY.
 FT SIGNAL. 20 707 LACTOTRANSFERRIN.
 RT CHAIN 20 357 1.
 FT REPEAT 358 707 2.
 FT DISULFID 27 63 BY SIMILARITY.
 FT DISULFID 37 54 BY SIMILARITY.
 FT DISULFID 133 216 BY SIMILARITY.
 FT DISULFID 175 191 BY SIMILARITY.
 FT DISULFID 188 199 BY SIMILARITY.
 FT DISULFID 249 263 BY SIMILARITY.
 FT DISULFID 366 398 BY SIMILARITY.
 FT DISULFID 376 389 BY SIMILARITY.
 FT DISULFID 423 702 BY SIMILARITY.
 FT DISULFID 443 665 BY SIMILARITY.
 FT DISULFID 475 550 BY SIMILARITY.
 FT DISULFID 499 693 BY SIMILARITY.
 FT DISULFID 509 523 BY SIMILARITY.
 FT DISULFID 520 533 BY SIMILARITY.
 FT DISULFID 591 605 BY SIMILARITY.
 FT DISULFID 643 648 BY SIMILARITY.
 FT METAL 78 78 IRON 1 (BY SIMILARITY).
 FT METAL 110 110 IRON 1 (BY SIMILARITY).
 FT METAL 210 210 IRON 1 (BY SIMILARITY).
 FT METAL 271 271 IRON 1 (BY SIMILARITY).
 FT METAL 413 413 IRON 2 (BY SIMILARITY).
 FT METAL 451 451 IRON 2 (BY SIMILARITY).
 FT METAL 544 544 IRON 2 (BY SIMILARITY).
 FT METAL 613 613 IRON 2 (BY SIMILARITY).
 FT BINDING 139 139 ANION (POTENTIAL).
 FT BINDING 481 481 ANION (POTENTIAL).
 FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1 2 MR -> IOG (IN REF. 1).
 FT CONFLICT 25 25 R -> Q (IN REF. 2).
 FT CONFLICT 82 82 M -> L (IN REF. 2).
 FT CONFLICT 359 359 S -> T (IN REF. 2).
 FT CONFLICT 382 382 A -> D (IN REF. 1).
 FT CONFLICT 449 449 E -> G (IN REF. 2).
 FT CONFLICT 629 629 L -> V (IN REF. 1).
 SQ SEQUENCE 707 AA; 77865 MW; F26AE0340A4C1948 CRC64;

Query Match 51.24; Score 84; DB 1; Length 707;
 Best Local Similarity. 44.44; Pred. No. 9.8e-05;
 Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 EWSKCYQWRKRLGAPSTICIRRTS 28
 DB 33 EBEKCLRMQNMKRVGGPPLSCVKKS 59

RESULT 7
 ID TRFL_HORSE STANDARD; PRT; 695 AA.
 AC 077811;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lactotransferrin precursor (lactoferrin) (Fragment).
 GN LTF.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Paramasivam M., Srinivasan A., Yadav M.P., Singh T.P.;
 RT "cDNA sequence of mare lactoferrin."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RC TISSUE=Milk;
 RX MEDLINE=99296631; Pubmed=10366507;

RA Sharma A.K., Paramasivam M., Srinivasan A., Yadav M.P., Singh T.P.;
 RT "Three-dimensional structure of mare deferic lactoferrin at 2.6-A
 resolution.";
 RL J. Mol. Biol. 289:303-317 (1999).
 CC -1- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ010930; CA09407.1; -.
 DR PDB; 1BIX; 02-DEC-98.
 DR PDB; 1B7U; 02-FEB-99.
 DR PDB; 1B7Z; 02-FEB-99.
 DR PDB; 1F9B; 10-FEB-01.
 DR PDB; 1IGB; 13-FEB-02.
 DR PDB; 1QJM; 14-JAN-00.
 DR InterPro; IPR01156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN 1; 2.
 DR PROSITE; PS00206; TRANSFERRIN 2; 2.
 DR PROSITE; PS00207; TRANSFERRIN 3; 1.
 DR Transprot; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal; 3D-structure.
 FT SIGNAL. 1 6
 FT CHAIN 7 695
 FT REPEAT 7 350
 FT DISULFID 15 51 LACTOTRANSFERRIN.
 FT DISULFID 25 42 1.
 FT DISULFID 121 204 2.
 FT DISULFID 163 179
 FT DISULFID 166 189
 FT DISULFID 176 187
 FT DISULFID 237 251
 FT DISULFID 354 386
 FT DISULFID 377 377
 FT DISULFID 411 690
 FT DISULFID 431 653
 FT DISULFID 463 538
 FT DISULFID 487 681
 FT DISULFID 497 511
 FT DISULFID 508 521
 FT DISULFID 579 593
 FT DISULFID 631 636
 FT METAL 66 66 IRON 1 (BY SIMILARITY).
 FT METAL 98 98 IRON 1 (BY SIMILARITY).
 FT METAL 198 198 IRON 1 (BY SIMILARITY).
 FT METAL 259 259 IRON 1 (BY SIMILARITY).
 FT METAL 401 401 IRON 1 (BY SIMILARITY).
 FT METAL 439 439 IRON 2 (BY SIMILARITY).
 FT METAL 532 532 IRON 2 (BY SIMILARITY).
 FT METAL 601 601 IRON 2 (BY SIMILARITY).
 FT BINDING 127 127 ANION (BY SIMILARITY).
 FT BINDING 469 469 ANION (BY SIMILARITY).
 FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 482 482 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT STRAND 12 16
 FT HELIX 19 34

```

FT TURN 35 36
FT STRAND 40 44
FT HELIX 48 56
FT TURN 57 58
FT STRAND 62 65
FT HELIX 67 74
FT TURN 76 78
FT STRAND 80 88
FT STRAND 95 95
FT STRAND 97 105
FT TURN 106 107
FT HELIX 112 114
FT TURN 116 117
FT STRAND 119 122
FT TURN 125 126
FT TURN 128 131
FT HELIX 132 137
FT HELIX 139 142
FT TURN 147 148
FT HELIX 151 158
FT STRAND 161 163
FT TURN 165 166
FT TURN 169 171
FT HELIX 173 175
FT TURN 176 178
FT TURN 183 187
FT TURN 191 192
FT TURN 194 195
FT HELIX 197 206
FT TURN 207 208
FT STRAND 212 216
FT HELIX 219 223
FT HELIX 227 230
FT TURN 231 232
FT STRAND 233 237
FT TURN 238 240
FT STRAND 241 244
FT HELIX 245 247
FT TURN 248 250
FT STRAND 254 257
FT STRAND 260 264
FT HELIX 270 284
FT TURN 286 287
FT TURN 294 295
FT TURN 308 309
FT STRAND 312 315
FT HELIX 322 326
FT TURN 328 336
FT HELIX 341 349
FT STRAND 351 356
FT HELIX 358 370
FT TURN 371 372
FT STRAND 375 380
FT HELIX 383 392
FT TURN 393 393
FT STRAND 397 400
FT HELIX 402 410
FT TURN 411 412
FT STRAND 414 421
FT HELIX 429 433
FT STRAND 439 446
FT HELIX 454 456
FT TURN 458 459
FT STRAND 462 464
FT TURN 467 468
FT TURN 470 473
FT HELIX 474 484
FT HELIX 489 491
FT TURN 492 492
FT STRAND 496 497
FT TURN 499 500
FT TURN 503 504

```

```

FT TURN 506 507
FT STRAND 508 508
FT TURN 515 516
FT TURN 520 521
FT TURN 525 526
FT TURN 528 529
FT HELIX 531 540
FT TURN 541 542
FT STRAND 546 550
FT HELIX 551 555
FT TURN 556 557
FT TURN 559 560
FT HELIX 565 568
FT TURN 569 572
FT TURN 572 573
FT STRAND 575 578
FT TURN 580 581
FT STRAND 584 586
FT HELIX 587 592
FT STRAND 596 598
FT STRAND 602 605
FT TURN 607 609
FT HELIX 610 624
FT TURN 626 627
FT TURN 629 630
FT HELIX 631 634
FT TURN 637 638
FT TURN 641 642
FT TURN 649 650
FT STRAND 651 655
FT TURN 658 659
FT HELIX 663 667
FT HELIX 669 679
FT TURN 680 681

```

Query Match 50.6%; Score 83; DB 1; Length 695;
 Best Local Similarity 50.0%; Pred. No. 0.00013;
 Matches 14; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

```

QY 2 EWSKCYQWQRKRGAPSTICIRRTSA 29
DB 21 EAAKCAKFORNMKVRGSPVSCIRKRTSS 48

```

```

RESULT 8
TRFL_PIG STANDARD; PRT; 704 AA.
AC P14632; Q29557;
DT 01-APR-1990 (Rel. 14, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactotransferrin precursor (Lactoferrin).
LTP.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RP [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92379101; PubMed=1511016;
RA Lyndon J.P., O'Malley B.R., Saucedo O., Lee T., Heaton D.R.,
RA Connely O.M.;
RT "Nucleotide and primary amino acid sequence of porcine lactoferrin.";
RL Biochim. Biophys. Acta 1132:97-99(1992).
RN [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=92367939; PubMed=1503259;
RA Alexander L.J., Levine W.B., Teng C.T., Beattie C.W.;
RT "Cloning and sequencing of the porcine lactoferrin cDNA.";
RL Anim. Genet. 23:251-256(1992).
RN [3]
RN SEQUENCE OF 20-49.
RP MEDLINE=90105538; PubMed=2605266;

```


DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 15-SEP-2003 (Rel. 42, last annotation update)
 DE Putative mannose-1,3-phosphoglycerate phosphatase (EC 3.1.3.70) (MGP)
 GN YGDP OR STM1986 OR STY2193 OR T0892.
 OS Salmonella typhimurium, and
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxId=602, 601;
 [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewl N., Mulvaney E.,
 RA Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhi; STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks R., Chillingworth T., Connerton P.,
 RA Cronin A., Davis P., Davies R.W., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
 RA Krogsh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18.";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- CATALYTIC ACTIVITY: 2(alpha-D-mannosyl)-3-phosphoglycerate + H(2)O.
 CC = 2(alpha-D-mannosyl)-D-glycerate + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HAD SUPERFAMILY. MGP FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE008788; AL20896.1; -
 DR EMBL; AL208788; CAD05733.1; -
 DR EMBL; AE016837; AA068570.1; -
 DR StvGene; SG77777; yedp.
 DR HAMAP; MF 00617; - 1.
 DR InterPro; IPR006379; HAD_SF_TTB.
 DR InterPro; IPR006381; HAD_SF_Yedp.
 DR InterPro; IPR005834; Hydrolase.
 DR Pfam; PF00702; Hydrolase; 1.
 DR TIGRfams; TIGR01486; HAD-SF-TTB-Yedp; 1.
 DR TIGRfams; TIGR01484; HAD-SF-TTB; 1.
 KM Hypothetical protein, Hydrolase; Complete proteome.
 SQ SEQUENCE 271 AA; 30925 MW; 46866C2E36B7008E CRC64;

Query Match 35.4%; Score 58; DB 1; Length 271;

Best Local Similarity 35.7%; Pred. No. 0.25;
 Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;
 Oy 2 EWSKCYQWRMRKXGAPSTCIRRTSA 29
 Db 25 EWOPAPWLRHESGVPLICSKTAA 52
 RESULT 13
 TRFM HUMAN STANDARD; PRT; 738 AA.
 AC P08582; Q9B0E2;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 15-SEP-2003 (Rel. 42, last annotation update)
 DE Melanotransferrin precursor (Melanoma-associated antigen p97) (CD228
 DE antigen).
 GN MF12 OR MAP97.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Melanoma;
 RX MEDLINE=86149285; PubMed=2419904;
 RA Rose T.M., Ploewman G.D., Teplov D.B., Dreyer W.J., Hellstrom K.E.,
 RA Brown J.P.;
 RT "Primary structure of the human melanoma-associated antigen p97
 RT (melanotransferrin) deduced from the mRNA sequence.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1261-1265(1986).
 [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Skin, and Uterus;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.U., Utsid T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullen S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek U.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Murray D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy U., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [3]
 RP GPI-ANCHOR.
 RX MEDLINE=94132080; PubMed=8300636;
 RA Food M.R., Rothenberger S., Gabathuler R., Haidl I.D., Reid G.,
 RA Jeffries W.A.;
 RT "Transposon and expression in human melanomas of a transferrin-like
 RT glycosylphosphatidylinositol-anchored protein.";
 RL J. Biol. Chem. 269:3034-3040(1994).
 [4]
 RP FUNCTION.
 RX MEDLINE=96016189; PubMed=7556058;
 RA Kennard M.L., Richardson D.R., Gabathuler R., Ponka P.,
 RA Jeffries W.A.;
 RT "A novel iron uptake mechanism mediated by GPI-anchored human p97.";
 RL EMBO J. 14:4178-4186(1995).
 [5]
 RP IRON-BINDING.
 RX MEDLINE=92183868; PubMed=1544447;

BA Baker E.N., Baker H.M., Smith C.A., Stebbins M.R., Kahn M.,
 RA Hellstrom K.E., Hellstrom I.,
 RA "Human melanotransferrin (p97) has only one functional iron-binding
 site.";
 RL FEBS Lett. 298:215-218(1992).
 RP [6]
 RN 3D-STRUCTURE MODELING.
 RX MEDLINE=92339524; PubMed=1633859;
 RA Garret R.C., Jhocti H.;
 RA "A molecular model for the tumour-associated antigen, p97, suggests a
 Zn-binding function";
 RL FEBS Lett. 305:55-61(1992).
 CC -1- FUNCTION: INVOLVED IN IRON CELLULAR UPTAKE. SEEMS TO BE
 CC INTERNALIZED AND THEN RECYCLED BACK TO THE CELL MEMBRANE. BINDS A
 CC SINGLE ATOM OF IRON PER SUBUNIT. COULD ALSO BIND ZINC.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=P08582-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P08582-2; Sequence=VSP_006557, VSP_006558;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: FOUND PREDOMINANTLY IN HUMAN MELANOMAS AND IN
 CC CERTAIN FETAL TISSUES; ALSO FOUND IN LIVER, EPITHELIUM, UMBILICAL
 CC CHORD, PLACENTA AND SWEAT GLAND DUCTS.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M12154; AAA05992.1; -;
 DR EMBL; A00127; CAA00012.1; -;
 DR EMBL; BC001875; AAH01875.1; -;
 DR EMBL; BC002623; AAH02623.1; -;
 DR EMBL; BC007550; AAH07550.1; -;
 DR FIR; A23814; TFPD.
 DR HSSP; P19134; 1TFD.
 DR GeneW; HGNC:7037; MF12.
 DR MIM; 155750; -;
 DR GO; GO:0005867; C: integral to plasma membrane; TAS.
 DR GO; GO:0008222; F: tumor antigen; TAS.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PRO0422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 DR TransPort; Iron transport; Glycoprotein; Metal-binding; Transmembrane;
 KM Repeat; Signal; GPI-anchor; Membrane; Zinc; Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 ? MELANOTRANSFERRIN.
 FT PROPEP 20 738 REMOVED IN MATURE FORM.
 FT REPEAT 20 361 1.
 FT DISULFID 362 713 2.
 FT DISULFID 26 63 BY SIMILARITY.
 FT DISULFID 130 54 BY SIMILARITY.
 FT DISULFID 172 189 BY SIMILARITY.
 FT DISULFID 186 199 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 FT METAL 78 78 IRON 1 (BY SIMILARITY).
 FT METAL 107 107 IRON 1 (BY SIMILARITY).
 FT METAL 210 210 IRON 1 (BY SIMILARITY).
 FT METAL 279 279 IRON 1 (BY SIMILARITY).
 FT METAL 421 421 IRON 2 (BY SIMILARITY).
 FT -----

FT METAL 451 451 IRON 2 (BY SIMILARITY).
 FT METAL 556 556 IRON 2 (BY SIMILARITY).
 FT METAL 625 625 IRON 2 (BY SIMILARITY).
 FT BINDING 136 136 ANION (POTENTIAL).
 FT TRANSMEM 714 738 ANCHOR.
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 238 302 GTPLEWGOALLSODELLCRGDSRADVTEMRCILARYPA
 FT HAVVPADPDGGLIPLINRGORT -> ESPSPROTWTSE
 FT EBBGCPAHEBARRTKRSAGQAMKAPVHRPODESDKGEF
 FT GRKSRDMLG (in isoform 2).
 FT /FTId=VSP_006557.
 FT Missing (in isoform 2).
 FT VARSPLIC 303 738 /FTId=VSP_006558.
 FT SO SEQUENCE 738 AA; 80241 MW; 6E6086E894D7B55 CRC64;
 Query Match 35.1%; Score 57.5; DB 1; Length 738;
 Best Local Similarity 43.3%; Pred. No. 0.85;
 Matches 13; Conservative 3; Mismatches 13; Indels 1; Gaps 1;
 QY 1 PEMSKCYQWQRMRKLG-APSITCIRRTSA 29
 DB 31 PEQHKCGNSEAFRENGIQPSLLCYRGISA 60
 RESULT 14
 ID PEMT_MOUSE STANDARD; PRT; 198 AA.
 AC Q61907;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE phosphatidy]ethanolamine N-methyltransferase (EC 2.1.1.17) (PEMT)
 DE (PEMT) (PEMT2).
 OS PEMT OR PEMT2.
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/J; TISSUE=Liver;
 RX MEDLINE=97133127; PubMed=8978486;
 RA Walkey C.J., Cui Z., Agellon L.B., Vance D.E.;
 RT "Characterization of the murine phosphatidy]ethanolamine
 RT N-methyltransferase-2 gene";
 RL J. Lipid Res. 37:2341-2350(1996).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=98058738; PubMed=9371769;
 RA Walkey C.J., Donohue L.R., Bronson R., Agellon L.B., Vance D.E.;
 RT "Disruption of the murine gene encoding phosphatidy]ethanolamine
 RT N-methyltransferase";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:12880-12885(1997).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98438461; PubMed=9765216;
 RA Walkey C.J., Yu L., Agellon L.B., Vance D.E.;
 RT "Biochemical and evolutionary significance of phospholipid
 RT methylation";
 RL J. Biol. Chem. 273:27043-27046(1998).
 CC -1- FUNCTION: CATALYZE THREE SEQUENTIAL METHYLATION OF
 CC PHOSPHATIDYLETHANOLAMINE (PE) BY ADOMET, THUS PRODUCING
 CC PHOSPHATIDYLCHOLINE (PC).
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine +
 CC phosphatidy]ethanolamine = S-adenosyl-L-homocysteine +
 CC phosphatidy]N-methyl ethanolamine.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL). FOUND
 CC IN MITOCHONDRIA AND IN ENDOPLASTIC RETICULUM.
 CC -1- TISSUE SPECIFICITY: LIVER.
 CC -1- SIMILARITY: BELONGS TO THE PEMT / PEM2 METHYLTRANSFERASE FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U25051; AAA67686.1; JOINED.
 DR EMBL; U25046; AAA67686.1; JOINED.
 DR EMBL; U25047; AAA67686.1; JOINED.
 DR EMBL; U25048; AAA67686.1; JOINED.
 DR EMBL; U25049; AAA67686.1; JOINED.
 DR EMBL; U25050; AAA67686.1; JOINED.
 DR MGD; MGI:104535; Pent.
 DR Pfam; PF04191; PENT; 1.

KW Phospholipid biosynthesis; Transferase; Methyltransferase;
 KW Transmembrane; Mitochondrion; Endoplasmic reticulum.
 FT INIT MET 0 0 BY SIMILARITY.
 FT TRANSMEM 12 32 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 157 177 POTENTIAL.
 SQ SEQUENCE 198 AA; 22385 MW; 5828975A55173D72 CRC64;

Query Match 31.7%; Score 52; DB 1; Length 198;
 Best Local Similarity 36.0%; Pred. No. 1.4;
 Matches 9; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

QY 3 WSKCYOMORRMRL---GAPSDTC 23
 Db 28 MNVAVRWEQRTRKLSRAFGSPHLAC 52

RESULT 15

MPGP_ECOL6 STANDARD; PRT; 271 AA.

AC P59286;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative mannosyl-3-phosphoglycerate phosphatase (EC 3.1.3.70) (MPGP).
 GN YEDP OR C2373.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;

SEQUENCE FROM N.A.

RC STRAIN=06:H1 / CFT073 / ATCC 700928;

RX MEDLINE=22388234; PubMed=12471157;

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,

RA Raeko D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,

RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,

RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;

RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).

RL -1- CATALYTIC ACTIVITY: 2(alpha-D-mannosyl)-3-phosphoglycerate + H(2)O

= 2(alpha-D-mannosyl)-D-glycerate + phosphate.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -1- SIMILARITY: BELONGS TO THE HAD SUPERFAMILY. MGP FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE016762; AAN80832.1; -

DR HAMAP; MF_00617; -; 1.

DR Pfam; PF00702; Hydrolase; 1.
 DR TIGRPFAM; TIGR01486; HAD-SF-11B-YedP; 1.
 DR TIGRPFAM; TIGR01484; HAD-SF-11B; 1.
 KW Hypothetical protein; Hydrolase; Complete proteome.
 SQ SEQUENCE 271 AA; 30515 MW; E6E78E3991470B2 CRC64;

Query Match 31.7%; Score 52; DB 1; Length 271;
 Best Local Similarity 32.1%; Pred. No. 2;
 Matches 9; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 2 EWSKCYOMORRMRLGAPSDTCIRRTSA 29
 Db 25 DWQPAAPWLSRLHEANIPVILCSKTS 52

Search completed: July 30, 2003, 16:24:30
 Job time : 10.494 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 30, 2003, 16:16:55 ; Search time 18.4702 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-5

Perfect score: 164
Sequence: 1 PEMSKCYQMQRMRKLGAPSTICIRRTSA 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	99.4	708	2 JC2323	lactoferrin - goat
2	122	74.4	708	1 TFBOL	lactotransferrin p
3	100	61.0	711	1 TFBOL	lactotransferrin p
4	91	55.5	33	2 S52107	lactoferrin - sheep
5	84	51.2	707	1 A28438	lactoferrin precursor
6	66	40.2	703	2 A45543	lactoferrin precursor
7	58.5	35.7	704	2 I47228	carbonic anhydrase
8	58	35.4	271	2 B85813	hypothetical prote
9	58	35.4	271	2 E90965	hypothetical prote
10	58	35.4	271	2 G64959	hypothetical prote
11	58	35.4	271	2 AB0754	conserved hypocher
12	57.5	35.1	738	1 TFBOL	melanotransferrin
13	51	31.1	199	2 A47353	phosphatidylethano
14	51	31.1	489	2 C70655	probable monooxyge
15	50	30.5	706	2 S33761	transferrin precursor
16	49	29.9	405	2 AB1461	B. subtilis yabE p
17	49	29.9	408	2 AC1098	B. subtilis yabE p
18	49	29.9	507	2 G87466	cryptophan halogen
19	49	29.9	1721	2 T21214	hypothetical prote
20	48.5	29.6	71	2 AH0033	hypothetical prote
21	48	29.3	430	2 D86573	histidyl tRNA synt
22	48	29.3	430	2 H72052	histidine-tRNA lig
23	47.5	29.0	423	2 F86313	F2H5.8 protein -
24	47	28.7	98	2 S39410	nosR protein - Par
25	47	28.7	528	2 T21834	hypothetical prote
26	47	28.7	795	2 T48252	eceriferum3 (CER3)
27	47	28.7	822	2 AB2507	hypothetical prote
28	47	28.7	2671	2 A49873	inositol 1,4,5-tri
29	46.5	28.4	121	2 AH3147	hypothetical prote

30	46.5	28.4	694	1 TFRBP	transferrin precursor
31	46	28.0	413	2 AH0780	probable oxidoredu
32	46	28.0	469	2 T26463	hypothetical prote
33	46	28.0	488	2 S36570	E2 protein - human
34	46	28.0	622	2 S61692	probable membrane
35	46	28.0	647	2 B8126	DNA topoisomerase
36	46	28.0	1082	2 T50650	AP-3 complex beta3
37	45.5	27.7	384	2 E69050	GlcNAc-phosphatidy
38	45.5	27.7	550	2 T47158	hypothetical prote
39	45.5	27.7	620	2 T27008	hypothetical prote
40	45.5	27.7	698	1 TFRBP	transferrin precursor
41	45	27.4	200	1 A42725	nitrite hydratase
42	45	27.4	284	2 S58650	hypothetical prote
43	45	27.4	310	2 C69218	transcription init
44	45	27.4	368	2 I55961	MHC class I histoc
45	45	27.4	395	2 T31578	hypothetical prote

ALIGNMENTS

RESULT 1
JC2323
lactoferrin - goat
C/Species: Capra aegagrus hircus (domestic goat)
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 07-May-1999
C/Accession: JC2323
R/Le Provost, F.; Nocard, M.; Guerin, G.; Martin, P.
Biochem. Biophys. Res. Commun. 203, 1324-1332, 1994
A/Title: Characterization of the goat lactoferrin cDNA: assignment of the relevant locus
A/Reference number: JC2323; UID:94380047; PMID:8093048
A/Accession: JC2323
A/Molecule type: mRNA
A/Residues: 1-708 <LEP>
C/Superfamily: transferrin; transferrin repeat homology
C/Keywords: duplication; glycoprotein
F/359-696/Domain: transferrin repeat homology <TRH2>
F/252,300,387,495,564/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match
Best Local Similarity 99.4%; Score 163; DB 2; Length 708;
Matches 28; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 33 PEMSKCYQMQRMRKLGAPSTICIRRTSA 29
|||||
1 PEMSKCYQMQRMRKLGAPSTICIRRTSA 29
|||||

RESULT 2
TFBOL
lactotransferrin precursor - bovine
N/Alternate names: lactoferrin
C/Species: Bos primigenius taurus (cattle)
C/Date: 31-Mar-1992 #sequence_revision 21-Nov-1997 #text_change 11-May-2000
C/Accession: I45919; S14674; S14110; J05095; S13097; S18518; S13881; P10148; S21
R/hsang, T.C.; Burns, D.K.; Wang, F.; Pan, Y.
FASEB J. 6, 233, 1991
A/Title: Cloning of a 80-kD advanced glycosylation end product (AGE) binding protein frc
A/Reference number: I45919
A/Accession: I45919
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-708 <TSA>
A/Cross-References: GB:I08604; NID:G163269; PTDN:AAA30609.1; PID:G163270
R/Pierce, A.
submitted to the EMBL Data Library, November 1990
A/Reference number: S14674
A/Accession: S14674
A/Molecule type: mRNA
A/Residues: 1-144, 'V', 146-163, 'PP', 166-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PI>
A/Cross-References: EMBL:X57084; NID:G505; PTDN:CAA40366.1; PID:G506
R/Pierce, A.; Colavizza, D.; Bernalsaez, M.; Maes, P.; Tartar, A.; Montreuil, J.; Spik, G.
Eur. J. Biochem. 196, 177-184, 1991

A>Title: Molecular cloning and sequence analysis of bovine lactoferrin.
 A:Reference number: S14110; MUID:91160550; PMID:2001696
 A:Accession: S14110
 A:Molecule type: mRNA
 A:Residues: 3-144, 'V', 146-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PI2>
 A:Cross-references: EMBL:X57084
 A:Accession: S18517
 A:Molecule type: protein
 A:Residues: 20-35,82-114,148-163, 'P', 166-178, 'V', 'P', 183-190,205-212,230-239,304-339,55
 R:Goodman, R.E.; Schanbacher, F.L.
 Biochem. Biophys. Res. Commun. 180, 75-84, 1991
 A>Title: Bovine lactoferrin mRNA: Sequence, analysis, and expression in the mammary gland
 A:Reference number: J70595; MUID:92028986; PMID:1718281
 A:Accession: J70595
 A:Molecule type: mRNA
 A:Residues: 1-65, 'P', 68-296, 'S', 298-339, 'A', 341-708 <GCO>
 A:Cross-references: GB:M63502
 A>Note: the authors translated the codon CCG for residue 66 as Arg and TCT for residue 2
 R:Meed, P.E.; Tweedle, J.W.
 Nucleic Acids Res. 18, 7167, 1990
 A>Title: cDNA and protein sequence of bovine lactoferrin.
 A:Reference number: S13097; MUID:91088328; PMID:2263492
 A:Accession: S13097
 A:Molecule type: mRNA
 A:Residues: 28-33, 'DS', 36-38, 'P', 40-708 <MEA>
 A:Cross-references: EMBL:X54801
 A:Accession: S18518
 A:Molecule type: protein
 A:Residues: 20-47,59-66,133-139,256-277,278,305-332,343-351,361-363,586,587-589,598-619
 R:Meed, P.E.
 Submitted to the EMBL Data Library, October 1990
 A:Reference number: S13881
 A:Accession: S13881
 A:Molecule type: mRNA
 A:Residues: 28-38, 'P', 40-86, 'C', 88-708 <ME3>
 A:Cross-references: EMBL:X54801
 R:Rejman, J.J.; Hegarty, H.M.; Hurley, M.L.
 Comp. Biochem. Physiol. B 93, 929-934, 1989
 A>Title: Purification and characterization of bovine lactoferrin from secretions of the
 A:Reference number: P10148; MUID:90031466; PMID:2805645
 A:Accession: P10148
 A:Molecule type: protein
 A:Residues: 20-27, 'X', 29-37, 'X', 39-54, 'X', 56-59 <RE3>
 R:Bellamy, W.; Takase, M.; Yamuchi, K.; Wakabayashi, H.; Kawase, K.; Tomita, M.
 Biochim. Biophys. Acta 1121, 130-136, 1992
 A>Title: Identification of the bactericidal domain of lactoferrin.
 A:Reference number: S21756; MUID:92287941; PMID:1599934
 A:Accession: S21756
 A:Molecule type: protein
 A:Residues: 36-60 <BL>
 R:Shimazaki, K.; Tanaka, T.; Kon, H.; Oota, K.; Kawaguchi, A.; Maki, Y.; Sato, T.
 J. Dairy Sci. 76, 946-955, 1993
 A>Title: Separation and characterization of the C-terminal half molecule of bovine lact
 A:Reference number: A56659; MUID:93253156; PMID:8486845
 A:Accession: A56659
 A:Molecule type: protein
 A:Residues: 20-25,302-308,359-366, 'X', 368-376, 'X', 378 <SHI>
 C:Superfamily: transferrin; lactoferrin repeat homology
 C:Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-708/Product: lactotransferrin #status experimental <MAT>
 F:20-355/Domain: transferrin repeat homology <TRH1>
 F:36-60/Region: antimicrobial
 F:359-60/Domain: transferrin repeat homology <TRH2>
 F:28-59,134-217,176-192,179-200,189-202,250-264,367-399,377-399,424-703,444-666,476-551,
 F:358-55/Denulfide bonds: #status predicted
 F:79,111,211,272/Binding site: iron (Asp, Tyr, Tyr, His) #status experimental
 F:140/Binding site: carbonate (Arg) #status experimental
 F:252,300,387,495,554/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:414,452,545,614/Binding site: iron (Asp, Tyr, Tyr, His) #status experimental
 F:482/Binding site: carbonate (Arg) #status experimental

Best Local Similarity 72.4%; Pred. No. 7.4e-10;
Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

RESULT 3
TEHUL

N; Alternate names: lactoferrin
C: Species: Homo sapiens (man)

C;Date: 31-Mar-1992 #sequence revision 21-Nov-1997 #text change 08-Dec-2000
C;Accession: G01394; S11228; A45401; S10324; S15853; S20841; S07160; A611694

R;Cho, Y.
submitted to the EMBL Data Library, March 1994

A/Accession: G01394
A/Accession number: 000020
A/Status: available
A/Status: available from CB/EMBL/DBP

A;Molecule type: mRNA
A;Residues: 1-711 <CHO>

A; Cross-references: EMBL:U07643; NID:g467236; PIDN:AA86
R; Rey, M.W.; Woloshuk, S.L.; deBoer, H.A.; Pieper, F.R.

A/Title: Complete nucleotide sequence of human mammary gland lactoferrin
A/Reference number: 511328, MIMD-80384829, DMTD-240245

A/Accession: S11228
A:Molecule type: mRNA

A/References: EMBL:X53961; NID:g34415; PIDN:CAA37914.1; PID:g34416

Mol. Endocrinol. 6, 1969-1981, 1992

A/Reference number: A45401; MUID:93125571; PMID:1480183
A/Accession: A45401

A/Residues: 1-15 <TEN>

A: Experimental source: placenta
A: Note: sequence extracted from NCBI backbone (NCBIP:122202)

R; Powell, M.J.; Ogden, J.E.
Nucleic Acids Res. 18, 4013, 1990

A:Reference number: S10324; MUID:90326549; PMID:2374734

A: Molecule type: mRNA
A: Residues: 3-711 <POW>

A; Cross-references: EMBL:X52941; NID:g34411; PIDN:CAAC3
R; Stowell, K.M.; Rado, T.A.; Funk, W.D.; Tweedie, J.W.

A/Title: Expression of cloned human lactoferrin in baby-hamster kidney cells
A/Reference number: 61663, MIM: 61364786, DMR: 204086

A:Accession: S15853
A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: mRNA
A;Residues: 20-31 <ST1>

A;Molecule type: protein
A;Residues: 20-28 /Y/ 30-31 -ST3-

R. Rado, T. A.; Wei, X.; Benz Jr., E. J.
Blood 70, 989-993, 1987

A: Reference number: S07160; MUID: 88001031; PMID: 3477300

A: Molecule type: mRNA
A: Residues: 436-487 / A' 489-711 (PAD)

A; Cross-references: EMBL:M18642; NID:g186815; PIR:R; Panella, T.J.; Liu, Y.; Huang, A.T.; Teng, C.T.

A: Title: Polymorphism and altered methylation of the lactoferrin gene in normal leukocytes
 Title: Polymorphism and altered methylation of the lactoferrin gene in normal leukocytes
 Cancer Res. 51, 3037-3043, 1991

A;Accession: A61169

Query Match 74.4%; Score 122; DB 1; Length 708

Query Match 40.2%; Score 66; DB 2; Length 703;
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
 C/Accession: E90965
 C/Species: *Escherichia coli*
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Gasaara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: E90965
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-271 <HAY>
 A/Cross-references: GB:BA000007; PIDN:BA36116.1; PID:G13362161; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain R1MD 0509952
 C/Genetics:
 A/Genes: Ec2693
 C/Superfamily: *Pyrococcus horikoshii* hypothetical protein PH0926

Query Match 40.2%; Score 66; DB 2; Length 703;
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
 C/Accession: E90965
 C/Species: *Escherichia coli*
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Gasaara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: E90965
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-271 <HAY>
 A/Cross-references: GB:BA000007; PIDN:BA36116.1; PID:G13362161; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain R1MD 0509952
 C/Genetics:
 A/Genes: Ec2693
 C/Superfamily: *Pyrococcus horikoshii* hypothetical protein PH0926

RESULT 7

147228
 carbonic anhydrase II inhibitor (transferrin homolog) precursor - pig
 C/Species: *Sus scrofa domestica* (domestic pig)
 C/Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 20-Aug-1999
 C/Accession: I47228
 R/Rouh, R.D.; Fietke, C.A.
 Biochemistry 31, 12536-12542, 1992
 A/Title: Purification and characterization of a carbonic anhydrase II inhibitor from por
 A/Reference number: I47228; MUID:93099129; PMID:1463741
 A/Accession: I47228
 A/Status: preliminary; translated from GB/EMBL/DBD
 A/Molecule type: mRNA
 A/Residues: 1-704 <RCU>
 A/Cross-references: EMBL:U36916; NID:G1016329; PIDN:AAB58956.1; PID:G1016330
 C/Genetics:
 A/Genes: PICA
 C/Superfamily: transferrin; transferrin repeat homology
 C/Keywords: duplication
 F/20-350/Domain: transferrin repeat homology <TRH1>

Query Match 35.7%; Score 58.5; DB 2; Length 704;
 Best Local Similarity 36.7%; Pred. No. 1.5;
 Matches 11; Conservative 8; Mismatches 8; Indels 3; Gaps 1;
 QY 2 EMSKCYQWQRRMKLGAPSTICIRRTS 28
 DB 34 EASKCSFRNMKKILPEGPHVSCVKTSS 63

RESULT 8

E85813
 hypothetical protein Z3045 (imported) - *Escherichia coli* (strain O157:H7, substrain EDL93
 C/Species: *Escherichia coli*
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Dec-2001
 C/Accession: B85813
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A/Reference number: A85480; MUID:21074935; PMID:11206551
 A/Accession: E85813
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-271 <STO>
 A/Cross-references: GB:AB005174; NID:91251601; PIDN:AAG56969.1; GSPDB:GN00145; UMGF:230
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Genes: Z3045
 C/Superfamily: *Pyrococcus horikoshii* hypothetical protein PH0926

Query Match 35.4%; Score 58; DB 2; Length 271;
 Best Local Similarity 35.7%; Pred. No. 0.71;
 Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 2 EMSKCYQWQRRMKLGAPSTICIRRTSA 29
 DB 25 DWQPAAPWLRRLREANVPVILCSKTS 52

RESULT 9

E90965
 hypothetical protein Ec2693 (imported) - *Escherichia coli* (strain O157:H7, substrain RJ

C/Species: *Escherichia coli*
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
 C/Accession: E90965
 C/Species: *Escherichia coli*
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Gasaara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: E90965
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-271 <HAY>
 A/Cross-references: GB:BA000007; PIDN:BA36116.1; PID:G13362161; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain R1MD 0509952
 C/Genetics:
 A/Genes: Ec2693
 C/Superfamily: *Pyrococcus horikoshii* hypothetical protein PH0926

Query Match 35.4%; Score 58; DB 2; Length 271;
 Best Local Similarity 35.7%; Pred. No. 0.71;
 Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

RESULT 10

G64959
 hypothetical protein b1955 - *Escherichia coli* (strain K-12)
 C/Species: *Escherichia coli*
 C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C/Accession: G64959
 R/Plattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of *Escherichia coli* K-12.
 A/Reference number: A64720; MUID:97426617; PMID:9278503
 A/Accession: G64959
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-271 <BLAT>
 A/Cross-references: GB:AE000287; GB:U00096; NID:G1788257; PIDN:AACT5021.1; PID:G1788265,
 A/Experimental source: strain K-12, substrain MG1655
 C/Superfamily: *Pyrococcus horikoshii* hypothetical protein PH0926

Query Match 35.4%; Score 58; DB 2; Length 271;
 Best Local Similarity 35.7%; Pred. No. 0.71;
 Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY 2 EMSKCYQWQRRMKLGAPSTICIRRTSA 29
 DB 25 DWQPAAPWLRRLREANVPVILCSKTS 52

RESULT 11

AB0754
 conserved hypothetical protein STY2193 (imported) - *Salmonella enterica* subsp. *enterica*
 C/Species: *Salmonella enterica* subsp. *enterica* serovar Typh
 A/Note: this species has also been called *Salmonella typhi*
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C/Accession: AB0754
 R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Conington, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
 A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero
 A/Reference number: AB0502; MUID:21534947; PMID:11677608
 A/Accession: AB0754
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-271 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05733.1; PID:gl6503226; GSPDB:GN00176
C:GeneticB:
A:Gene: STY2193
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0926

Query Match	35.4%	Score	58	DB	2	Length	271
Best Local Similarity	35.7%	Pred	NO	0.71			
Matches	10	Conservative	4	Mismatches	14	Indels	0
						Gaps	0

```
QY      2 EWSKCYQWQRMRKLGAPISITCIR TSA 23
          || | : | | | : |
Db      25 EWQPAFPWLTRLAESGVPIILCSSKTAA 52
```

RESULT 12

melanotransferrin precursor - human
N; Alternate names: melanoma-associated antigen gp95/p97

A;Gene: GDB:MF12
A;Cross-references: GDB:119387; OMIM:155750
A;Map position: 3q28-3q29

Query Match	35.1%	Score 57.5	DB 1	Length 738
Best Local Similarity	43.3%	Pred. No. 2.2		
Matches 13; Conservative	3	Mismatches 13	Indels 1	Gaps 1

```
OY      1 PMSKCYQWQRMRKLG-APSITCIRRTSA 29
          ||| | : | | : | : |||
DB     31 PEQHKGNSSEAFREAGIQPSLLCVRTISA 60
```

RESULT 13

phosphatidylethanolamine N-methyltransferase (EC 2.1.1.17), 20K - rat
C_Species: Rattus norvegicus (Norway rat)

A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-199 <CUT>
A/Cross-references: GB:U14441, NID:G310194, PIDN:AAA03154.1, PID:G310195
A/Superfamily: phosphatidy1-N-methyl ethanolamine N-methyltransferase
C/Keywords: membrane protein; methyltransferase; S-adenosylmethionine

Query Match	31.1%;	Score 51;	DB 2;	Length 199;
Best Local Similarity	36.0%;	Pred. No. 5.6;		
Matches	9;	Conservative	6;	Mismatches 4;
				Gaps 1

```

QY      3  MSKCYQWQRRMRKL---GAPSI TC 23
          | : : : : | | | | : | : |
Db.     29  MNVVARWEQRTKL SRAFGSPYIAC 53

```

RESULT 14

Probable monooxygenase - C ₁ species: Mycobacterium	Mycobacterium tuberculosis (strain H37RV)
tuberculosis	

A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: RV3854c

```
QY      3 WSKCYQWQRRMRL 16
        :|:|:|:|:|:
Db      250 YSACQKWPRRMRLM 26
```

RESULT 15

transferrin precursor - horse
N; Alternate names: growth-promoting factor

F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-706/Product: transferrin #status experimental <MAT>
F;358-694/Domain: transferrin repeat homology <TRH2>
F;26-64,36-55,134-215,174-190,177-198,187-200,248-262,360-623,366-398,376-389,423-701,44

Query Match 30.5%; Score 50; DB 2; Length 706;

Best Local Similarity 34.5%; Pred. No. 26;
Matches 10; Conservative 7; Mismatches 10; Indels 2; Gaps 1;

QY 2 EWSKCYQWQRMRKL--GAPSTICIRRTS 28

Db 32 EWSKASFRDMSKSIIVAPPLVACVKRTS 60

Search completed: July 30, 2003, 16:31:17
Job time : 20.4702 secs

DR WPI, 2000-271377/23.
XX
XX Novel process for producing peptides with e.g. antimicrobial activity
PT from biological fluids such as milk, whey or blood comprises contacting
PT fluid with chromatographic medium to adsorb peptide domain of interest
PT
PS Claim 14, Page 22; 41pp; English.
XX
XX This sequence represents a peptide derived from an internal fragment of
CC goat whey. The peptide is an example of a peptide with antibacterial
CC activity that can be produced by the process of the invention. The
CC invention relates to a process for producing peptides from biological
CC fluids. The process comprises chromatography of the biological fluid, in
CC situ hydrolysis of selectively bound peptides, washing to remove unbound
CC peptide, and elution of the peptides of interest. The process is used for
CC producing peptides from biological fluids, such as milk, whey or blood.
CC For example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumour activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptides with a
CC selected activity of interest without the need for intermediate
CC purification of the precursor protein.
XX
XX Sequence 29 AA;
SQ
Query Match 100.0%; Score 164; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 8.4e-17;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PEMSKCYQWQRMRKLGAPSTICIRRTSA 29
DB 1 PEMSKCYQWQRMRKLGAPSTICIRRTSA 29
RESULT 2
AAG80774 ID AAG80774 standard; Protein; 708 AA.
XX
XX AAG80774;
AC
XX 19-APR-2002 (first entry)
DT
XX Goat lactoferrin-associated protein #1.
DE
XX
XX Primer; lactoferrin; antibiotic; goat.
KM
XX
XX Capra sp.
OS
XX
XX Key Location/Qualifiers
FH MISC-difference 360 /note= "Encoded by CCG"
FT
XX
XX KR98043944-A.
PN
XX
XX 05-SEP-1998.
PD
XX
XX 05-DEC-1996; 96KR-0061925.
PF
XX
XX 05-DEC-1996; 96KR-0061925.
PR
XX
XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.
PA
XX
XX Yoo DY, Nam MS, Lee TH, Yoo SR, Kim SJ, Lee GG;
PI
XX
XX WPI; 1999-465631/39.
DR
XX
XX N-PSDB; ABA97187.
DR
XX
XX Antibiotic lactoferrin derived from Korean black goat and gene thereof -
PT
XX
XX NoAbstract

PS Disclosure; Page 6-9; 11pp; Korean.
XX
XX This sequence represents the goat antibiotic lactoferrin described in the
CC disclosure of the invention.
CC
XX
XX Sequence 708 AA;
SQ
Query Match 100.0%; Score 163; DB 20; Length 708;
Best Local Similarity 96.6%; Pred. No. 2.6e-15;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PEMSKCYQWQRMRKLGAPSTICIRRTSA 29
DB 33 PEMSKCYQWQRMRKLGAPSTICIRRTSA 61
RESULT 3
AAG80775 ID AAG80775 standard; Protein; 708 AA.
XX
XX AAG80775;
AC
XX 19-APR-2002 (first entry)
DT
XX
XX Goat lactoferrin-associated protein #2.
DE
XX
XX Primer; lactoferrin; antibiotic; goat.
KM
XX
XX Capra sp.
OS
XX
XX Key Location/Qualifiers
FH MISC-difference 360 /note= "Encoded by CCG"
FT
XX
XX KR98043944-A.
PN
XX
XX 05-SEP-1998.
PD
XX
XX 05-DEC-1996; 96KR-0061925.
PF
XX
XX 05-DEC-1996; 96KR-0061925.
PR
XX
XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.
PA
XX
XX Yoo DY, Nam MS, Lee TH, Yoo SR, Kim SJ, Lee GG;
PI
XX
XX WPI; 1999-465631/39.
DR
XX
XX Antibiotic lactoferrin derived from Korean black goat and gene thereof -
PT
XX
XX NoAbstract
PS Disclosure; Page 6-9; 11pp; Korean.
XX
XX This sequence represents the goat antibiotic lactoferrin described in the
CC disclosure of the invention.
CC
XX
XX Sequence 708 AA;
SQ
Query Match 99.4%; Score 163; DB 20; Length 708;
Best Local Similarity 96.6%; Pred. No. 3.6e-15;
Matches 28; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 PEMSKCYQWQRMRKLGAPSTICIRRTSA 29
DB 33 PEMSKCYQWQRMRKLGAPSTICIRRTSA 61
RESULT 4
AAV83112 ID AAV83112 standard; peptide; 25 AA.
XX
XX
XX AAV83112;
AC
XX
XX

DT 24-JUL-2000 (first entry)
 DE Synthetic lactoferrin fragment.
 XX
 XX Lactoferrin; antimicrobial; cytostatic; infection; tumour;
 KM sterilisation; wound healing; spermicide; goat.
 XX
 OS Synthetic.
 OS Capra hircus.
 XX
 PN MO200012541-A2.
 PD
 XX 09-MAR-2000.
 PF 31-AUG-1999; 99WO-GB02850.
 XX
 PR 26-AUG-1998; 98GB-0018938.
 XX
 PA (ALPH-) ALPHARMA AS.
 PA (GARD/) GARDNER R.
 XX
 PI Svendsen JS, Rekdal O, Sveinbjornsson B, Vorland L;
 DR MPI, 2000-270793/23.
 XX
 PT Cytotoxic peptides useful as medicament for treating tumors and
 PT bacterial infections, comprises one or more non-genetic bulky and
 PT lipophilic amino acids
 XX
 PS Example 1; Figure 1; 114pp; English.
 XX
 CC Cytotoxic 7-25 mer lactoferrin peptides which comprise three or more
 CC cationic residues and have one or more non-genetic bulky and
 CC lipophilic amino acids have cytostatic and antimicrobial activity.
 CC The peptides are useful as medicament for treating bacterial
 CC infections and tumours. They are also useful for sterilising
 CC agents for materials susceptible to microbial contamination. They may
 CC also be used as promoters of wound healing and spermicides. The small
 CC size of the peptides gives them an increased half life and allows
 CC efficient biodelivery. The peptides can be administered without need
 CC for an injection, such as by inhalation or by absorption across the
 CC blood capillaries of the nasal passages. This synthetic lactoferrin
 CC peptide corresponds to amino acids 17-41 of caprine lactoferrin.
 XX
 SQ Sequence 25 AA;
 Query Match 82.9%; Score 136; DB 21; Length 25;
 Best Local Similarity 96.0%; Pred. No. 8.8e-13;
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 4 SKCYQWQRMRKLGAPSTICRRTS 28
 DB 1 SKCYQWQRMRKLGAPSTICRRTS 25
 ID AAY85013 standard; peptide; 25 AA.
 AC AAY85013;
 XX
 DT 20-JUN-2000 (first entry)
 XX
 DE Caprine lactoferrin peptide LFC(17-41) amino acid sequence.
 KM Lactoferrin; lactoferricin; antibacterial; bacterial infection;
 KM bacterial growth; solid tumour; treatment; goat.
 XX
 OS Capra sp.
 OS
 PN MO200012542-A2.
 XX
 PD 09-MAR-2000.
 XX

XX
 PF 31-AUG-1999; 99WO-GB02851.
 XX
 PR 26-AUG-1998; 98GB-0018938.
 XX
 PA (ALPH-) ALPHARMA AS.
 PA (GARD/) GARDNER R.
 XX
 PI Svendsen JS, Rekdal O, Sveinbjornsson B, Vorland L;
 DR MPI, 2000-256582/22.
 XX
 PT Modified cytotoxic lactoferrin peptide useful for preparing
 PT pharmaceutical composition against bacterial infections and tumours and
 PT for inhibiting bacterial growth
 XX
 PS Example 1; Fig 1; 112pp; English.
 XX
 CC This sequence represents a caprine lactoferricin peptide. Lactoferrin is
 CC a weak iron transporter which also shows weak antibacterial effect.
 CC Lactoferricin B (LFB) is a peptide produced when bovine lactoferrin is
 CC digested with pepsin. LFB is more active than lactoferrin. The two
 CC tryptophan residues at positions 6 and 8 in LFB 17-31 cannot be
 CC substituted with alanine without loss of antibacterial activity. The
 CC invention relates to a modified cytotoxic lactoferrin peptide of 7-25
 CC amino acids in length with three cationic residues and one or more extra
 CC bulky and lipophilic amino acids, than its native peptide. The invention
 CC also relates to the esters, amides, salts and cyclic derivatives of the
 CC modified peptide. The modified lactoferrin peptides of the invention are
 CC useful for treating bacterial infections and tumours, for preparing
 CC medicaments and for inhibiting bacterial growth. LFB and its fragments
 CC are useful for manufacturing medicaments for treating solid tumours. The
 CC modified peptides have a very small size, which is suitable for
 CC biodelivery. Due to the smaller size, the circulating half-life of the
 CC peptide is increased as they are less vulnerable to endopeptidases. The
 CC observed cytolytic effect of the modified lactoferrin peptides in tumours
 CC is not species specific and they have greater utility in treating human
 CC tumours.
 XX
 SQ Sequence 25 AA;
 Query Match 82.9%; Score 136; DB 21; Length 25;
 Best Local Similarity 96.0%; Pred. No. 8.8e-13;
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 4 SKCYQWQRMRKLGAPSTICRRTS 28
 DB 1 SKCYQWQRMRKLGAPSTICRRTS 25
 ID AAY49270 standard; protein; 689 AA.
 AC AAY49270;
 XX
 DT 07-FEB-2000 (first entry)
 XX
 DE Lactoferrin sequence.
 KM Adhesion inhibition; enteropathogenic; Escherichia coli; lactoferrin;
 KM medicament; drink; feed; food poisoning.
 XX
 OS Unidentified.
 OS
 PN JP11292789-A.
 XX
 PD 26-OCT-1999.
 XX
 PF 03-APR-1998; 98JP-0107167.
 XX
 PR 03-APR-1998; 98JP-0107167.
 XX

PA (MEIP) MEIJI MILK PROD CO LTD.
 XX
 DR WPI; 2000-018674/02.
 XX
 PT Adhesion inhibitory compositions of enteropathogenic E. coli to cells -
 XX used in medicines and foodstuffs
 XX
 XX Disclosure; Fig 4; 7pp; Japanese.
 XX
 CC The invention provides an adhesion inhibitory composition of
 CC enteropathogenic Escherichia coli to cells. The composition contains a
 CC lactoferrin degraded matter. The composition is useful as a medicament,
 CC a drink and food or a feed. The composition is effective for prevention
 CC and/or treatment of infectious food poisoning caused by enteropathogenic
 CC E. coli.
 CC
 XX
 SQ Sequence 689 AA;
 Query Match 74.4%; Score 122; DB 21; Length 689;
 Best Local Similarity 72.4%; Pred. No. 3.4e-09;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 Oy 1 PEMSKCYQWRMRKLGAPSTICIRRTSA 29
 Db 14 PEMFKCRWQMRMKLGAPSTICVRRAFA 42
 RESULT 7
 AAM09343
 ID AAM09343 standard; Protein; 708 AA.
 AC AAM09343;
 XX
 DT 25-MAR-2003 (updated)
 DT 18-MAR-1997 (first entry)
 XX
 DE Bovine lactoferrin.
 XX
 KM Human; lactoferrin; iron-binding glycoprotein; milk; secretion; fungus;
 KM transferin; bactericidal activity; prostate; expression system; primer;
 KM PCR; polymerase chain reaction; amplification; signal peptide; antiviral;
 KM alpha-amylase; Aspergillus oryzae; nutrition; bovine.
 OS
 OS Bos taurus.
 XX
 PN USS571691-A.
 XX
 PD 05-NOV-1996.
 XX
 PF 28-OCT-1993; 93US-0145681.
 XX
 PR 28-OCT-1993; 93US-0145681.
 PR 05-MAY-1989; 89US-0348270.
 PR 28-SEP-1989; 89US-0413860.
 PR 24-APR-1992; 92US-0873304.
 PR 27-OCT-1992; 92US-0967947.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX
 PI Conneely OM, Heaton DR, May GS, O'Malley BW;
 XX
 DR WPI; 1996-505400/50.
 DR N-PSDB; AAT48031.
 XX
 PT New DNA encoding human lactoferrin, its natural alleles and
 PT substitution analogues - useful e.g. for preventing iron deficiency
 PT and as antiviral/antimicrobial agent
 XX
 XX Disclosure; Column 31-34; 92pp; English.
 XX
 CC This is the amino acid sequence of bovine lactoferrin, an iron-binding
 CC glycoprotein found in milk and other secretions and body fluids.
 CC Fragments of the protein are also known to have biological activity e.g.

CC the N-terminal portion of the protein has a bactericidal activity. The
 CC gene was used to construct a fusion protein in which the native
 CC lactoferrin signal peptide was replaced by the alpha-amylase II signal
 CC peptide. The novel construct was then expressed in Aspergillus oryzae.
 CC The protein can be used for antibacterial and antiviral activities as
 CC well as an iron-carrying protein for nutritional or therapeutic
 CC applications.
 CC (Updated on 25-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 708 AA;
 Query Match 74.4%; Score 122; DB 17; Length 708;
 Best Local Similarity 72.4%; Pred. No. 3.5e-09;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 Oy 1 PEMSKCYQWRMRKLGAPSTICIRRTSA 29
 Db 33 PEMFKCRWQMRMKLGAPSTICVRRAFA 61
 RESULT 8
 AAM57318
 ID AAM57318 standard; Protein; 708 AA.
 AC AAM57318;
 XX
 DT 10-AUG-1998 (first entry)
 XX
 DE Bovine lactoferrin.
 XX
 KM Bovine; lactoferrin; recombinant; therapeutic; nutritional; iron;
 KM Fe binding site; bacteria; bactericidal; milk.
 XX
 OS
 OS Bos taurus.
 XX
 PN USS766939-A.
 XX
 PD 16-JUN-1998.
 XX
 PF 30-MAY-1995; 95US-0453703.
 XX
 PR 28-OCT-1993; 93US-0145681.
 PR 05-MAY-1989; 89US-0348270.
 PR 24-APR-1992; 92US-0873304.
 PR 27-OCT-1992; 92US-0967947.
 PR 30-MAY-1995; 95US-0453703.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX
 PI Conneely OM, Heaton DR, May GS, O'Malley BW;
 XX
 DR WPI; 1998-361705/31.
 DR N-PSDB; AAV30771.
 XX
 PT Plasmids containing human lactoferrin DNA - for recombinant
 PT production of the enzyme, especially fragments having bactericidal
 PT activity
 XX
 XX Example 13; Fig 14; 92pp; English.
 XX
 CC The present sequence represents bovine lactoferrin. A plasmid has been
 CC developed which is suitable for the expression of a human lactoferrin,
 CC or an iron-binding lobe of lactoferrin, in a transformed prokaryotic
 CC host cell. The plasmid comprises a DNA sequence encoding a naturally
 CC occurring human lactoferrin protein or an iron-binding lobe of
 CC lactoferrin and further comprises transcriptional and translational
 CC regulatory elements capable of regulating the expression of the
 CC lactoferrin-encoding DNA sequence in the transformed host cell. The
 CC plasmid is useful for producing recombinant human lactoferrin proteins
 CC in bacteria. Lactoferrin is a 78 kDa iron-binding glycoprotein found
 CC in milk and other secretory fluids. It is involved in iron transfer
 CC and delivery in mammals. It has been implicated as a resistance
 CC factor in suckled new born infants against enteritis infections; the

CC resistance to bacterial and viral infection. The tissue specific
 CC expression of human lactoferrin in mammary glands, for instance,
 CC impacts the bactericidal and virucidal benefit of the expressed gene to
 CC young feeding on the milk and provides a production means for the
 CC secreted protein for therapeutic use.

XX Sequence 708 AA;

Query Match 74.4%; Score 122; DB 21; Length 708;
 Best Local Similarity 72.4%; Pred. No. 3.5e-09;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 PEWSKYQWQRRMKLGAPISITCIRRTSA 29
 DB 33 PEWFKCRWQRRMKLGAPISITCVRRAFA 61

RESULT 11

AG64827 standard; protein; 708 AA.

XX AG64827;

DT 20-SEP-2001 (first entry)

DE Chronic hepatitis treatment related protein SEQ ID NO: 8.

XX Chronic hepatitis; viral antigenic protein; hepatitis C; hepatitis B.

OS Bos taurus.

PN W0200147545-A1.

PD 05-JUL-2001.

PF 28-DEC-2000; 2000WO-JP09393.

PR 28-DEC-1999; 99JP-0374087.

PA (SUMU) SUMITOMO PHARM CO LTD.

PI Tohdoh N, Murata M, Enjoji T;

DR WPI; 2001-425585/45.

PT Treatment and prevention of chronic hepatitis

XX Example 1; Page 74-78; 128pp; Japanese.

CC The present invention describes a method of preventing and treating

CC chronic hepatitis, involving administering an oligopeptide which

CC (a) has binding affinity towards the viral antigenic protein;

CC (b) inhibits binding affinity of the virus towards the receptor protein

CC of the target cell and

CC (c) has analogy with the receptor protein at the amino acid level. This

CC can be used to prevent and treat hepatitis B and C. The present sequence

CC is a protein described in the exemplification of the invention.

XX Sequence 708 AA;

QY 1 PEWSKYQWQRRMKLGAPISITCIRRTSA 29

DB 33 PEWFKCRWQRRMKLGAPISITCVRRAFA 61

RESULT 12
 ID AAE02342 standard; Protein; 708 AA.
 XX

AC AAE02342;

DT 10-AUG-2001 (first entry)

DE Bovine lactoferrin (LF).

XX Bovine; lactoferrin; LF; therapeutic; nutritional; iron transport;

XX virucidal; bactericidal; animal food; iron-binding glycoprotein.

OS Bos taurus.

PN US6228614-B1.

PD 08-MAY-2001.

PF 09-MAR-1999; 99US-0265577.

PR 28-OCT-1993; 93US-0145681.

PR 05-MAY-1989; 89US-0348270.

PR 30-MAY-1995; 95US-0456108.

PR 24-APR-1992; 92US-0873304.

PR 27-OCT-1992; 92US-0967947.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Conneely OM, Headon DR, O'malley BW, May GS;

XX WPI; 2001-342673/36.

XX N-PSDB; AAD06282.

XX Producing recombinant human lactoferrin useful for therapeutic or

XX nutritional applications, comprises transforming a eukaryotic cells

XX with a vector having a DNA that encodes and permits expression of

XX lactoferrin in the cells

XX Example 13; Fig 14B; 92pp; English.

XX The present invention relates to a method for expressing human

XX lactoferrin in an eukaryotic cell. The method comprises introducing

XX a plasmid comprising DNA encoding a human lactoferrin protein and

XX regulatory elements necessary for the expression of DNA in the cell.

XX The method is used for preparing recombinant lactoferrin, making

XX applications, e.g. as therapeutic additives to enhance iron transport

XX and delivery and for virucidal or bactericidal qualities or as human

XX or animal food. The present method is an efficient and economical way

XX to produce human lactoferrin.

XX The present sequence is bovine lactoferrin (LF). LF is an

XX iron-binding glycoprotein found in milk, other secretions and body

XX fluids.

XX Sequence 708 AA;

QY 1 PEWSKYQWQRRMKLGAPISITCIRRTSA 29

DB 33 PEWFKCRWQRRMKLGAPISITCVRRAFA 61

Query Match 74.4%; Score 122; DB 22; Length 708;
 Best Local Similarity 72.4%; Pred. No. 3.5e-09;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

RESULT 13
 ID AAR44845 standard; peptide; 32 AA.
 XX AAR44845;
 XX

DT 25-MAR-2003 (updated)
 DT 09-JAN-2003 (updated)
 DT 31-JAN-1994 (first entry)
 XX Lactoferrin-related antibacterial peptide.
 DE

XX Lactoferrin; antibiotic; chelate; mastitis; bowel disorder;
 KW disease; bacteria; yeast; fungi; disinfection; drug; foodstuff;
 KW cosmetic; toiletries.
 XX
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT Disulfide bond 10..27
 XX
 PN MO9314640-A1.
 PD
 XX 05-AUG-1993.
 PD
 PF 30-NOV-1992; 92WO-JP01563.
 XX
 PR 23-JUN-1992; 92JP-0032660.
 PR 11-MAR-1992; 92JP-0052943.
 PR 30-SEP-1992; 92JP-0262143.
 PR 30-SEP-1992; 92JP-0262559.
 XX
 PA (MORG) MORINAGA MILK IND CO LTD.
 PI Bellamy W, Fukuwatari Y, Kawase K, Shimamura S;
 PI Takase M, Tokitay, Tomita M, Wakabayashi H, Yamauchi K;
 DR WPI; 1993-258265/32.
 XX
 PT Antibacterial agent comprising decomposition products of
 PT lactoferrin - with chelate e.g. EDTA alcohol and/or antibiotic
 PT e.g. penicillin, also useful against yeast and fungi
 XX
 PS Disclosure; Page 87; 100pp; Japanese.
 XX
 CC Lactoferrin-related peptides are used in new antibacterial compn.
 CC The compn. is highly effective against a broad range of bacteria,
 CC yeasts and fungi. It can be used therapeutically (internal and
 CC external application), e.g. for mastitis, bowel disorders, urinary
 CC infections, etc. It can also be used for the disinfection and
 CC protection of drugs, foodstuffs, cosmetics and toiletries and
 CC household items (such as kitchen towels and toilet paper).
 CC (Updated on 09-JAN-2003 to add missing OS field.)
 CC
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 SQ Sequence 32 AA;
 XX
 Query Match 73.2%; Score 120; DB 14; Length 32;
 Best Local Similarity 76.9%; Pred. No. 2.5e-10;
 Matches 20; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 PEMSKCYQMORRMKLGAPSTICRR 26
 DB 5 PEMFKCRMQMWMKMLGAPSTICVRR 30
 XX
 RESULT 14
 AAR48534
 ID AAR48534 standard; peptide; 32 AA.
 XX
 AC AAR48534;
 XX
 DT 25-MAR-2003 (updated)
 DT 10-AUG-1994 (first entry)
 XX
 DE Lactoferrin derived peptide #28.
 XX
 KW Decomposition; lactoferrin; digestion; enzyme; pepsin; trypsin;
 KW antioxidant; oxidation; inhibitor; vitamin E; ascorbic acid;
 KM vitamin A; beta-carotene; superoxide dismutase; coenzyme Q;
 KM lipid oxidation; foodstuff; drugs; health food; toiletries; cosmetics.
 XX
 OS Bos taurus.
 XX

PN WO9403555-A1.
 XX
 PD 17-FEB-1994.
 XX
 PF 04-AUG-1993; 93WO-JP01090.
 XX
 PR 07-AUG-1992; 92JP-0211335.
 XX
 PA (MORG) MORINAGA MILK IND CO LTD.
 PI Bellamy WR, Fukuwatari Y, Kawase K, Shimamura S;
 PI Takase M, Tokitay, Tomita M, Wakabayashi H, Yamauchi K;
 DR WPI; 1994-065650/08.
 XX
 PT Antioxidant peptide lactoferrin decomposition product - prevents
 PT oxidation of lipid(s) in foodstuffs and drugs without affecting
 PT their taste
 XX
 PS Claim 3; Page 39; 47pp; Japanese.
 XX
 CC The sequences given in AAR48507-37 are peptides derived by the
 CC decomposition of lactoferrin, pref. by digestion with an enzyme, eg.
 CC pepsin or trypsin. These peptides may be used in an antioxidant
 CC composition which may also contain an oxidation inhibitor such as
 CC vitamin E, ascorbic acid, vitamin A, beta-carotene, superoxide
 CC dismutase or coenzyme Q. The antioxidant prevents lipid oxidation
 CC in foodstuffs, drugs, health foods, toiletries and cosmetics.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 SQ Sequence 32 AA;
 XX
 Query Match 73.2%; Score 120; DB 15; Length 32;
 Best Local Similarity 76.9%; Pred. No. 2.5e-10;
 Matches 20; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 PEMSKCYQMORRMKLGAPSTICRR 26
 DB 5 PEMFKCRMQMWMKMLGAPSTICVRR 30
 XX
 RESULT 15
 AAR57466
 ID AAR57466 standard; Protein; 32 AA.
 XX
 AC AAR57466;
 XX
 DT 28-FEB-1995 (first entry)
 XX
 DE Lactoferrin derived peptide #29.
 XX
 KW Lactoferrin; chemical; enzymatic; hydrolysis; antimicrobial;
 KW antiseptic; ischaemic disease.
 XX
 OS Mus musculus.
 XX
 PN JP06172200-A.
 XX
 PD 21-JUN-1994.
 XX
 PF 08-DEC-1992; 92JP-0327738.
 XX
 PR 08-DEC-1992; 92JP-0327738.
 XX
 PA (MORG) MORINAGA MILK IND CO LTD.
 PI WPI; 1994-238662/29.
 XX
 DR Brain protectant for preventing ischaemic diseases without side
 PT effects - comprising 31 specified peptide(s), prepd. by
 PT lactoferrin hydrolysis
 XX
 PS Disclosure; Page 10; 11pp; Japanese.
 XX

XX The sequences given in AAR57438-68 represent fragments of lactoferrin
CC which were derived from the full length protein by chemical or enzyme
CC hydrolysis. These peptides have brain protecting properties, as
CC well as anti-microbial activity. Compositions containing these
CC peptides may be prepared with out the addition of antiseptics, and
CC may be administered at doses of at least 10 mg for parenteral
CC administration and 100 mg for oral administration. These peptides
CC are stable, heat resistant, water soluble and may be used for the
CC prevention of ischaemic diseases without side effects.

XX
SQ Sequence 32 AA:

Query Match 73.2%; Score 120; DB 15; Length 32;
Best Local Similarity 76.9%; Pred. No. 2.5e-10;

Matches 20; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 PEMSKCYQWQRRMKLGAPSTICRR 26
||| | : ||| | ||| | ||| | ||| |

DB 5 PEMFKCRWQWRMKLGAPSTICRR 30
||| | : ||| | ||| | ||| | ||| |

Search completed: July 30, 2003, 16:23:28
Job time : 58.1369 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 49.1664 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-6

Perfect score: 164
Sequence: 1 PEMSKCYQWQRMKRGAPSTICIRRTSA 29

Scoring table: BLOSUM62

Searched: 830525 seqs, 258052604 residues
Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertibrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	61.0	234	4	Q8IX02
2	99	60.4	711	4	Q8TCD2
3	97	59.1	711	4	Q8I2H6
4	97	59.1	711	4	Q8I092
5	91	55.5	33	6	Q9TR80
6	84	51.2	707	11	Q8CB40
7	69	42.1	704	6	Q8WNR8
8	68	41.5	38	4	Q9UCY5
9	60.5	36.9	700	11	Q9DBD0
10	58	35.4	711	6	Q9XT72
11	57.5	35.1	700	11	Q8VC96
12	54.5	33.2	192	2	Q8KL39
13	54	32.9	821	2	Q8KRP9
14	53.5	32.6	736	6	Q97490
15	53	32.3	48	6	Q77558
16	53	32.3	1361	12	Q8BF56

17	53	32.3	1449	12	Q65974	Q65974 caesava com
18	52	31.7	199	11	Q8R011	Q8R011 mus musculu
19	51.5	31.4	1795	4	Q14674	Q14674 homo sapien
20	51.5	31.4	4437	2	Q8GGP3	Q8GGP3 streptomyc
21	51	31.1	489	16	P96223	P96223 mycobacteri
22	50	30.5	54	6	Q9TQV7	Q9TQV7 equus cabal
23	50	30.5	154	16	Q8PF84	Q8PF84 xanthomonas
24	50	30.5	511	11	Q8BVH0	Q8BVH0 mus musculu
25	50	30.5	522	5	Q9VWL4	Q9VWL4 drosophila
26	50	30.5	643	11	Q8BV16	Q8BV16 mus musculu
27	50	30.5	1207	4	Q8TB59	Q8TB59 homo sapien
28	49.5	30.2	1336	4	Q9C014	Q9C014 homo sapien
29	49	29.9	130	17	Q8TVE4	Q8TVE4 methanopyru
30	49	29.9	195	12	Q9IWM2	Q9IWM2 hydrangea r
31	49	29.9	258	5	Q9VAF6	Q9VAF6 drosophila
32	49	29.9	405	16	Q92F81	Q92F81 listeria in
33	49	29.9	408	16	Q8YAB4	Q8YAB4 listeria mo
34	49	29.9	507	16	Q9A7G8	Q9A7G8 caulobacter
35	49	29.9	558	2	Q936J3	Q936J3 rhodospirill
36	49	29.9	1721	5	Q93547	Q93547 caenorhabdi
37	48.5	29.6	71	16	Q8ZJ54	Q8ZJ54 yersinia pe
38	48.5	29.6	132	16	Q8CLO9	Q8CLO9 yersinia pe
39	48.5	29.6	518	16	Q9L1H4	Q9L1H4 streptomyc
40	48	29.3	271	16	Q98JP8	Q98JP8 rhizobium i
41	48	29.3	352	7	Q9XS13	Q9XS13 aulonocara
42	48	29.3	362	16	Q8XQ97	Q8XQ97 raietonia s
43	48	29.3	393	10	Q9ZTP0	Q9ZTP0 oryza sativ
44	48	29.3	402	10	Q9ZRH8	Q9ZRH8 oryza sativ
45	48	29.3	424	16	Q98FV3	Q98FV3 rhizobium i

ALIGNMENTS

RESULT 1	
Q8IX02	PRELIMINARY; PRT; 234 AA.
ID Q8IX02	
AC Q8IX02	
DT 01-MAR-2003 (TREMREL. 23, Created)	
DT 01-MAR-2003 (TREMREL. 23, Last sequence update)	
DE Lactoferrin (Fragment).	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RA Liu Z., Li N.,	
RT "Human lactoferrin gene, 5' regulation region and exons 1-6.";	
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.	
DR EMBL; AF508798; AAC14686.1; -.	
FT	
FT NON_TER	
SQ	
SEQUENCE 234 AA; 25671 MW; 2A08DAFDB8780C7 CRC64;	
Query Match	61.0%; Score 100; DB 4; Length 234;
Best Local Similarity	57.1%; Pred. No. 66-07; Mismatches 6; Indels 0; Gaps 0;
Matches 16; Conservative	
QY	
1 PEMSKCYQWQRMKRGAPSTICIRRTS 28	
: : : :	
DB	
34 PEATKCFQWQRMKRGAPSPVSCIKRDS 61	
RESULT 2	
Q8TCD2	PRELIMINARY; PRT; 711 AA.
ID Q8TCD2	
AC Q8TCD2	
DT 01-JUN-2002 (TREMREL. 21, Created)	
DT 01-JUN-2002 (TREMREL. 21, Last sequence update)	
DE Lactoferrin.	
OS Homo sapiens (Human).	

RESULT	6
Q8CBA0	
Q8CBA0	PRELIMINARY; PRT; 707 AA.
AC	Q8CBA0;
DT	01-MAR-2003 (TrEMBLrel. 23, Created)
DT	01-MAR-2003 (TrEMBLrel. 23, last sequence update)
DT	01-MAR-2003 (TrEMBLrel. 23, last annotation update)
DE	Lactotransferrin.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Bone;
RX MEDLINE=22354683; PubMed=12466851;
RA THE FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573(2002).
DR EMBL; AK036491; BAC29450.1; -.
SQ SEQUENCE 707 AA; 77837 MW; E1B32F5FD8748A0F CRC64;

Query Match 51.2%; Score 84; DB 11; Length 707;
Best Local Similarity 44.4%; Pred. No. 0.00044;
Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 EWSKCYQWRMRKLGAPSLTCIRRTS 28
DB 33 EEERKLRWQEMRKVGPPPLSCVKSS 59
|||:|||||:|||||:|||||:
RESULT 7
ID Q8WMN8 PRELIMINARY; PRT; 704 AA.
AC Q8WMN8;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Lactoferrin (Fragment).
GN LTF.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
CX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Wang S.-R., Lin T.-Y., Weng C.-N.;
RT "Isolation and expression of porcine milk lactoferrin.";
RL Submitted (DEC-2001) to the EMBL/Genbank/DDBP databases.
CC -!- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC CC EMBL; L77887; AAL40161.1; -.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN 1; 2.
DR PROSITE; PS00206; TRANSFERRIN 2; 2.
DR PROSITE; PS00207; TRANSFERRIN 3; 1.
DR Glycoprotein; Iron transport; Metal-binding; Transport.
KW NON TER
FT 1
SQ SEQUENCE 704 AA; 77681 MW; 64EE769F7503CC32 CRC64;

Query Match 42.1%; Score 69; DB 6; Length 704;
Best Local Similarity 51.9%; Pred. No. 0.077;
Matches 14; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY 2 EWSKCYQWRMRKLGAPSLTCIRRTS 28
DB 34 EYSKCRQWQSKIRRTN--PIFCIRAS 58
|||:|||||:|||||:
RESULT 8
ID Q9UCY5 PRELIMINARY; PRT; 38 AA.
AC Q9UCY5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

```

Query Match	Best Local Similarity	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
1	PEMSKCYQWQRMRKLGAPSTTCI 24	41.5%	68	38	12	4	8	0	0
15	PEADKXFGQMRNMRKXRGVXSXI 38	50.0%	Preq. No. 0.006						

ID	Q9DBD0	PRELIMINARY;	PRT;	700 AA.
AC	Q9DBD0			
DT	01-JUN-2001 (TrEMBLrel. 17, Created)			
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	1300017J02RIK protein.			
GN	1300017J02RIK			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Liver;			
RC	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukumitsu Y., Konno H., Adachi J., Fukuda S.,			
RA	Aitawa K., Iwawa M., Nishi K., Kiyosawa H., Kondo S., Yamazaki R.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nkaido I., Pesole G., Quackenbush J.,			
RA	Schirini L.M., Staehli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., But C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Guatinicich S., Hill D., Hofmann M., Hume D.A., Kamlay M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,			
RA	Hayashizaki Y.;			
RT	"Functional annotation of a full-length mouse cDNA collection."			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK005035; BAB23762.1; -			
DR	HSSP; P19134; ITPD.			
DR	MGD; MGI:1919025; 1300017J02RIK.			
DR	InterPro; IPR001156; Transferrin.			
DR	Pfam; PF00405; Transferrin_2.			
DR	PRINTS; PR00422; TRANSFERRIN.			
DR	SMART; SMO0094; TR_FER; 2.			
DR	PROSITE; PS00205; TRANSFERRIN_1; 1.			
DR	PROSITE; PS00207; TRANSFERRIN_3; 1.			

```

SQ  SEQUENCE 700 AA; 76765 MW; 38C991D1021A548 CRC64;

Query Match          36.9%; Score 60.5; DB 11; Length 700;
Best Local Similarity 36.7%; Pred. No. 1.4;
Matches 11; Conservative 9; Mismatches 7; Indels 3; Gaps 1.

OY  2 EWSKCYOWQRMKRL--GAPSTTCIRRTS 28
    | :|:|:|:|:|:|:|:|:|:|:|:|:|
DB  34 EATKCSFRDNKKVLPAGGPAVTCVRKMS 63

RESULT 10
O9XT72 PRELIMINARY; PRT; 711 AA.
AC O9XT72.
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Serum iron transport protein transferrin.
GN TRFE.
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
OX NCBI_TaxId=9337;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA MEDLINE=99227120; PubMed=10209259;
RA Demmer J., Stasiuk S.J., Adamski F.M., Grigor M.R.;
RT "Cloning and expression of the transferrin and ferritin genes in a
RT marsupial, the brushtail possum (Trichosurus vulpecula).";
RL Biochim. Biophys. Acta 1445:65-74(1999).
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC CC
DR DR EMBL; AF092510; AAD38331.1; -.
DR DR HSBP; P56410; IAOV.
DR DR InterPro; IPR001156; Transferrin.
DR DR Pfam; PF00405; transferrin; 2.
DR DR PRINTS; PR00422; TRANSFERRIN.
DR DR SMART; SMO0094; TR_PFR; 2.
DR DR PROSITE; PS00206; TRANSFERRIN_2; 1.
DR DR PROSITE; PS00207; TRANSFERRIN_3; 1.
DR DR GlycoProtex; Iron transport; Metal-binding; Transport.
SQ SEQUENCE 711 AA; 77764 MW; 35CA4594DD95631F CRC64;

Query Match          35.4%; Score 58; DB 6; Length 711;
Best Local Similarity 33.3%; Pred. No. 3.4;
Matches 9; Conservative 7; Mismatches 11; Indels 0; Gaps 0.

OY  2 EWSKCYOWQRMKRLGAPSTTCIRRTS 28
    | :|:|:|:|:|:|:|:|:|:|:|:|:|
DB  33 ELKCNFRKMNNSINHGGLACTIKTTS 59

RESULT 11
O8VC96 PRELIMINARY; PRT; 700 AA.
AC O8VC96.
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE RIKEN cDNA 130001.7J02 gene.
GN CN
GN 130001.7J02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strassberg R.;

```

```

DL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC021390.1; AAB1390.1; -
DR MGD; MG1:1918025; 130001702Rik.
DR InterPro; IPR01156; transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PRO0422; TRANSFERRIN.
DR SMART; SMO0094; TR_FER_2.
DR PROSITE; PS00205; TRANSFERRIN_1; 1.
DR PROSITE; PS00207; TRANSFERRIN_3; 1.
SQ SEQUENCE 700 AA; 76795 MW; 7DA4A580CB6BF37 CRC64;

Query Match      35.1%; Score 57.5; DB 1; Length 700;
Best Local Similarity 37.0%; Pred. No. 4;
Matches 10; Conservative 8; Mismatches 6; Indels 3; Gaps 1;

QY          5 KCYQMQRMRKTL--GAPSTICIRRTS 28
           |||::|||::|||::|||::|||
Db         37 KCSFSFDNMKKVLPAGGPATVCVAKMS 63

RESULT 12
Q8KL39      PRELIMINARY; PRT; 192 AA.
ID   Q8KL39
AC   Q8KL39;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein.
GN   yp044.
OS   Rhizobium etl.
OC   Plasmid symbiotic plasmid p42d.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=29449;
[1]
RN     SEQUENCE FROM N.A.
RP     STRAIN=CFM42;
RC     MEDLINE=91193195; PubMed=2013564;
RA     Girard M.L., Flores M., Brom S., Romero D., Palacios R., Davila G.;
RT     "Structural complexity of the symbiotic plasmid of Rhizobium
RL     leguminosarum bv. phaseoli."
RL     J Bacteriol. 173:2411-2419(1991).
[2]
RN     SEQUENCE FROM N.A.
RP     STRAIN=CFM42;
RC     MEDLINE=97419521; PubMed=9274036;
RA     Ramirez-Romero M.A., Bustos P., Girard L., Rodriguez O.,
RA     Cevallos M.A., Davila G.;
RT     "Sequence, localization and characteristics of the replicator region
RL     of the symbiotic plasmid of Rhizobium etl.";
RL     Microbiology 143:2825-2831(1997).
[3]
RN     SEQUENCE FROM N.A.
RP     STRAIN=CFM42;
RA     Quinteiro V., Cevallos M.A., Davila G.;
RT     "A site-specific recombinase and RecA are required to exert
RL     incompatibility towards the symbiotic plasmid of Rhizobium etl.";
RL     Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80928; AAM54928.1; -
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE 192 AA; 21513 MW; 7AC7624DF0EB232E CRC64;

Query Match      33.2%; Score 54.5; DB 2; Length 192;
Best Local Similarity 60.0%; Pred. No. 3.1;
Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY          9 WQRRMKLAGPSTICIRRTS 28
           |||::|||::|||::|||::|||
Db        148 WQERGSGAP-VTAIRRNS 166

RESULT 13
08KPR9

```

ID 08KPR9 PRELIMINARY; PRT; 821 AA.
 AC 08KPR9;
 DT 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE Hypothetical protein.
 GN SEB0023.
 OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
 OX NCBI_TaxId=1140;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC 7942;
 RA Holman C.K., Sandoval P., Chen Y., Socias T., Mohler B.J.,
 RA McMurtry S., Gonzalez A., Salinas I., Golden S.S., Youderlian P.;
 RT "Synecococcus elongatus PCC7942 cosmid 7G3";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF120853; JAM82700.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 821 AA; 91706 MW; D5221A0D0ED0DBCC CRC64;

Query Match 32.9%; Score 54; DB 2; Length 821;
 Best Local Similarity 47.8%; Pred. No. 16;
 Matches 11; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Qy 1 PMSKCYQWRMRKLGAPSITC 23
 Db 517 PQSQAEIWLQWRALTLAKPITC 539

RESULT 14
 O97490 PRELIMINARY; PRT; 736 AA.
 AC 097490;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE P97.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxId=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98451505; PubMed=9780225;
 RA Kawamoto T., Pan H., Yan W., Ishida H., Uenu E., Oda R., Nakamasu K.,
 RA Noshiro M., Kawashima-Ohya Y., Fujii M., Shintani H., Okada Y.,
 RA Kato Y.;
 RT "Expression of membrane-bound transferrin-like protein p97 on the cell
 RT surface of chondrocyte";
 RL Eur. J. Biochem. 256:503-509(1998).
 CC -!- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC EMBL; AB010995; BAA33956.1; -.
 DR HSSP; P19134; ITFD.
 DR InterPro; IPR001064; Cysteallin.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; Transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 1.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.
 DR Glycoprotein; Itron transport; Metal-binding; Transport.
 SQ SEQUENCE 736 AA; 80169 MW; F389D9F8A6AC90FC CRC64;

Query Match 32.6%; Score 53.5; DB 6; Length 736;
 Best Local Similarity 36.7%; Pred. No. 17;
 Matches 11; Conservative 6; Mismatches 12; Indels 1; Gaps 1;

Qy 1 PMSKCYQWRMRKLG-APSITCIRRTSA 29
 Db 31 PQQCEWDSQAFRAGLQPALCVQGTSA 60

RESULT 15
 O77558 PRELIMINARY; PRT; 48 AA.
 AC 077558;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Lactoferrin (Fragment).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 OX NCBI_TaxId=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98332734; PubMed=9666128;
 RA Wang S.R., Lin J., Cheng I.C., Lin T.Y.;
 RT "Characterization and functional analysis of the porcine lactoferrin
 RT gene promoter";
 RL Gene 215:203-212(1998).
 DR EMBL; AF044256; AAC34369.1; -.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 1.
 FT NON TER 48 48
 SQ SEQUENCE 48 AA; 5569 MW; 28403BE7DE14AD78 CRC64;

Query Match 32.3%; Score 53; DB 6; Length 48;
 Best Local Similarity 57.1%; Pred. No. 1.3;
 Matches 8; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Search completed: July 30, 2003, 16:29:25
 Job time : 49.1964 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 ; Search time 9.49405 Seconds

(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-6

Perfect score: 164
Sequence: 1 PMSKCYQWQRRMRKLGAPSLTCIRRTSA 29

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	164	100.0	708	1 TRFL_CAPHI	Q29477 capra hircu
2	124	75.6	708	1 TRFL_BOBBU	O77698 bubalus bub
3	122	74.4	708	1 TRFL_BOVIN	P24627 bos taurus
4	104	63.4	708	1 TRFL_CAMDR	O91000 camelus dro
5	100	61.0	711	1 TRFL_HUMAN	P02788 homo sapien
6	84	51.2	707	1 TRFL_MOUSE	P08071 mus musculu
7	83	50.6	695	1 TRFL_HORSE	O77811 equus cabal
8	66	40.2	704	1 TRFL_PIG	P14632 sus scrofa
9	58.5	35.7	704	1 ICA_PIG	O29545 sus scrofa
10	58	35.4	271	1 MRGP_ECO57	O82039 escherichia
11	58	35.4	271	1 MRGP_ECOLI	P76329 escherichia
12	58	35.4	271	1 MRGP_SALTY	O82039 escherichia
13	57.5	35.1	738	1 TRFL_HUMAN	P08582 homo sapien
14	52	31.7	198	1 PENT_MOUSE	O61907 mus musculu
15	52	31.4	271	1 MRGP_ECOLI	P59286 escherichia
16	51.5	31.4	1795	1 ESP1_HUMAN	Q14674 homo sapien
17	51	31.1	198	1 PENT_RAT	O08388 rattus norv
18	50	30.5	642	1 Z398_HUMAN	O8217 homo sapien
19	50	30.5	706	1 TRFE_HORSE	P27425 equus cabal
20	50	30.5	1207	1 AT19_HUMAN	O81659 homo sapien
21	48	29.3	430	1 SYH_CHLUP	O92721 chlamydia p
22	47.5	29.0	450	1 FEM2_RAT	O92721 chlamydia p
23	47	28.7	198	1 PENT_RAT	O92721 chlamydia p
24	47	28.7	704	1 TRFL_HUMAN	O92721 chlamydia p
25	47	28.7	704	1 TRFL_HUMAN	O92721 chlamydia p
26	46.5	28.4	2671	1 IP3T_HUMAN	O12931 homo sapien
27	46	28.0	695	1 TRFE_RABIT	Q14573 homo sapien
28	46	28.0	63	1 BD03_MOUSE	O92721 chlamydia p
29	46	28.0	413	1 YEIT_SALTY	O82566 salmonella
30	46	28.0	488	1 VET_SALTY	O82566 salmonella
31	46	28.0	647	1 TOP3_HPV49	P36795 human papil
32	46	28.0	1082	1 A3B2_HUMAN	O92721 chlamydia p
33	46	28.0	1210	1 AT19_MOUSE	P59509 mus musculu

34	45.5	27.7	641	1 SCAB_RABIT	O97742 oryctolagus
35	45.5	27.7	698	1 TRFE_HUMAN	P02787 homo sapien
36	45.5	27.7	967	1 AT51_HUMAN	O92721 chlamydia p
37	45	27.4	199	1 NHAA_PSECL	P27764 pseudomonas
38	45	27.4	310	1 TRFB_METHH	O26971 methanobact
39	45	27.4	368	1 HA14_MOUSE	P14427 mus musculu
40	45	27.4	1385	1 RPOD_PLAVV	O07518 plantago as
41	45	27.4	1490	1 CRK7_HUMAN	O92721 chlamydia p
42	44.5	27.1	937	1 SYL_METHH	O27552 methanobact
43	44.5	27.1	1793	1 VCP1_LOTIA	O92721 chlamydia p
44	44	26.8	142	1 RPOD_ORNEB	P31843 oenothera b
45	44	26.8	236	1 YH1_ECOLI	P32133 escherichia

ALIGNMENTS

RESULT 1
ID TRFL_CAPHI STANDARD; PRT; 708 AA.
AC Q29477; Q29479;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactoferrin precursor (Lactoferrin).
GN LTF.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Lee T., Yu S., Kim S., Lee K., Yu D.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA MEDLINE=94380047; PubMed=8093048;
RT "Characterization of the goat lactoferrin cDNA. Assignment of the
RT relevant locus to bovine U12 syntenic group."
RT Biochem. Biophys. Res. Commun. 203:1324-1332(1994).
CC -!- FUNCTION: TRANSFERRIN ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U53857; AAA97958.1; -
CC EMBL; X78902; CA55517.1; -
CC HSSP; O77698; 1CER;
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; Transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR_FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.

```

FT CHAIN 20 708 LACTOTRANSFERRIN.
FT REPEAT 20 708 1.
FT REPEAT 364 708 2.
FT DISULFID 28 64 BY SIMILARITY.
FT DISULFID 38 55 BY SIMILARITY.
FT DISULFID 134 217 BY SIMILARITY.
FT DISULFID 176 192 BY SIMILARITY.
FT DISULFID 189 200 BY SIMILARITY.
FT DISULFID 250 264 BY SIMILARITY.
FT DISULFID 367 399 BY SIMILARITY.
FT DISULFID 377 390 BY SIMILARITY.
FT DISULFID 424 703 BY SIMILARITY.
FT DISULFID 444 666 BY SIMILARITY.
FT DISULFID 476 551 BY SIMILARITY.
FT DISULFID 500 694 BY SIMILARITY.
FT DISULFID 510 524 BY SIMILARITY.
FT DISULFID 521 534 BY SIMILARITY.
FT DISULFID 592 606 BY SIMILARITY.
FT DISULFID 644 649 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 211 211 IRON 1 (BY SIMILARITY).
FT METAL 272 272 IRON 1 (BY SIMILARITY).
FT METAL 414 414 IRON 2 (BY SIMILARITY).
FT METAL 452 452 IRON 2 (BY SIMILARITY).
FT METAL 545 545 IRON 2 (BY SIMILARITY).
FT METAL 614 614 IRON 2 (BY SIMILARITY).
FT BINDING 140 140 ANION (BY SIMILARITY).
FT BINDING 482 482 ANION (BY SIMILARITY).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 56 56 I -> V (IN REF. 2).
FT CONFLICT 88 88 L -> R (IN REF. 2).
FT CONFLICT 124 124 Q -> K (IN REF. 2).
FT CONFLICT 154 154 F -> P (IN REF. 2).
FT CONFLICT 304 304 S -> R (IN REF. 2).
FT CONFLICT 414 414 D -> G (IN REF. 2).
SQ SEQUENCE 708 AA; 77358 MW; P2EDA3C83539960D CRC64;

```

Query Match 100.0%; Score 164; DB 1; Length 708;
 Best Local Similarity 100.0%; Pred. No. 1.4e-16;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 PMSKCYQWQRRMKLGAPSLTIRRTSA 29
DB 33 PMSKCYQWQRRMKLGAPSLTIRRTSA 61

```

```

RESULT 2
TRFL_BUBBU STANDARD; PRT; 708 AA.
ID TRFL_BUBBU
AC 077698;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lactotransferrin precursor (lactoferrin).
GN LTF.
OS Bubalus bubalis (Domestic water buffalo).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bubalus.
OX NCBI_TaxID=89462;
RN [1]
RP SEQUENCE FROM N.A.
RA Paramesivam M., Thattaiyath B.D., Kumar A., Srinivasan A.,
RT "cDNA sequence of Buffalo lactoferrin";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

```

```

RX MEDLINE=20003130; PubMed=10531476;
RA Karthikeyan S., Paramesivam M., Yadav S., Srinivasan A., Singh T.P.;
RT "Structure of buffalo lactoferrin at 2.5-A resolution using crystals
RT grown at 303 K shows different orientations of the N and C lobes.";
RL Acta Crystallogr. D 55:1805-1813(1999).
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A005203; CA06441.1; -.
DR PDB; 1CR2; 19-MAR-99.
DR PDB; 1BY1; 13-JAN-99.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SM00094; TR FER; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
DR Transprot; Iron transport; Glycoprotein; Metal-binding; Repeat;
KM Signal; 3d-structure.
FT CHAIN 1 19
FT DISULFID 20 708
FT DISULFID 28 64
FT DISULFID 38 55
FT DISULFID 134 217
FT DISULFID 176 192
FT DISULFID 179 202
FT DISULFID 189 200
FT DISULFID 250 252
FT DISULFID 250 252
FT DISULFID 367 399
FT DISULFID 377 390
FT DISULFID 424 703
FT DISULFID 444 666
FT DISULFID 476 551
FT DISULFID 500 694
FT DISULFID 510 524
FT DISULFID 521 534
FT DISULFID 592 606
FT DISULFID 644 649
FT DISULFID 79 79
FT METAL 111 111 IRON 1.
FT METAL 211 211 IRON 1.
FT METAL 272 272 IRON 1.
FT METAL 414 414 IRON 2.
FT METAL 452 452 IRON 2.
FT METAL 545 545 IRON 2.
FT METAL 614 614 IRON 2.
FT BINDING 140 140 ANION (POTENTIAL).
FT BINDING 482 482 ANION (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TURN 21 22
FT STRAND 25 29
FT HELIX 32 46
FT TURN 47 48
FT STRAND 53 57
FT HELIX 61 69
FT TURN 70 71

```

```

FT STRAND 75 78
FT HELIX 80 87
FT TURN 89 91
FT STRAND 93 102
FT STRAND 107 108
FT STRAND 110 118
FT TURN 119 120
FT HELIX 125 127
FT TURN 129 130
FT STRAND 132 135
FT TURN 138 139
FT TURN 141 144
FT HELIX 145 150
FT TURN 151 151
FT HELIX 152 155
FT TURN 159 161
FT HELIX 164 169
FT TURN 170 171
FT STRAND 174 176
FT TURN 178 179
FT TURN 182 184
FT HELIX 186 189
FT TURN 190 191
FT HELIX 196 198
FT TURN 199 200
FT TURN 204 205
FT TURN 207 208
FT HELIX 211 219
FT TURN 220 221
FT STRAND 225 229
FT TURN 230 231
FT HELIX 232 236
FT HELIX 240 243
FT TURN 244 245
FT STRAND 246 249
FT TURN 251 252
FT STRAND 255 257
FT HELIX 258 260
FT TURN 261 263
FT STRAND 267 270
FT STRAND 273 277
FT HELIX 283 297
FT TURN 299 300
FT TURN 307 308
FT TURN 312 313
FT TURN 321 322
FT STRAND 325 328
FT TURN 331 332
FT HELIX 335 351
FT HELIX 354 362
FT STRAND 364 369
FT HELIX 371 384
FT TURN 385 388
FT STRAND 388 393
FT HELIX 396 404
FT TURN 405 406
FT STRAND 410 413
FT HELIX 415 422
FT TURN 423 425
FT STRAND 426 434
FT TURN 438 441
FT TURN 444 446
FT STRAND 452 459
FT TURN 460 461
FT TURN 463 464
FT HELIX 467 469
FT TURN 471 472
FT STRAND 474 477
FT TURN 480 481
FT TURN 483 486
FT HELIX 487 497
FT TURN 503 505
FT STRAND 508 510

```

```

FT TURN 512 513
FT TURN 516 517
FT TURN 519 520
FT TURN 522 523
FT TURN 528 529
FT TURN 533 534
FT TURN 538 539
FT TURN 541 542
FT HELIX 544 553
FT TURN 554 555
FT STRAND 559 563
FT HELIX 564 569
FT TURN 570 570
FT HELIX 578 581
FT TURN 582 582
FT TURN 585 586
FT STRAND 588 591
FT TURN 593 594
FT STRAND 597 599
FT TURN 600 601
FT TURN 603 605
FT STRAND 609 611
FT STRAND 615 619
FT HELIX 620 637
FT TURN 639 640
FT TURN 642 647
FT TURN 650 651
FT TURN 654 655
FT TURN 662 663
FT STRAND 664 669
FT HELIX 676 680
FT HELIX 682 692
FT TURN 693 694
FT HELIX 698 705

```

Query Match
 Best Local Similarity 75.6%; Score 124; DB 1; Length 708;
 Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 PEMSKCYOWMRKLGAPSTICRR 26
 DB 33 PEMLKCHRWQMRKLGAPSTICVAR 58

RESULT 3
 ID TRFL_BOVIN STANDARD; PRT; 708 AA.
 AC P24637; Q29629; Q9MZY3;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Lactotransferrin precursor (Lactoferrin) [Contains: Lactoferricin B (LFCIN B)].
 DE LTF.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Submaxillary gland;
 RX MEDLINE=91160550; PubMed=2001696;
 RA Pierce A., Colavizza D., Benaisa M., Maes P., Tartar A.,
 RA Montreuil J., Spik G.;
 RL "Molecular cloning and sequence analysis of bovine lactotransferrin.";
 RL Eur. J. Biochem. 196;177-184(1991).
 RP [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92028986; PubMed=1718281;
 RA Goodman R.E., Schanbacher F.L.;
 RT "Bovine lactoferrin mRNA: sequence, analysis, and expression in the
 RT mammary gland.";

RL Biochem. Biophys. Res. Commun. 180:75-84(1991).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Teag T.C., Burns D.K., Wang F., Pan Y.C.E., Schmidt A.M., Stern D.;
 RT "Cloning of a 80-kD advanced glycosylation end product (AGE) binding
 RL protein from bovine lung."
 RL FASEB J. 6:233-233 (1991).
 RN (4)
 RN SEQUENCE FROM N.A.
 RC TISSUE=Blood, and Mammary gland;
 RX MEDLINE=94266164; PubMed=8206385;
 RA Seyfert H.-M., Tuckovic A., Interthal H., Koczan D., Hobom G.;
 RL "Structure of the bovine lactoferrin-encoding gene and its promoter."
 RN Gene 143:265-269(1994).
 RN (5)
 RN SEQUENCE FROM N.A.
 RA Nakamura I., Shimazaki K., Yagi Y., Watanabe A.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN SEQUENCE OF 20-59.
 RX MEDLINE=90031466; PubMed=2805645;
 RA Rejman J.J., Hegarty H.M., Hurley M.L.;
 RT "Purification and characterization of bovine lactoferrin from
 RL secretions of the involuting mammary gland: identification of
 multiple molecular weight forms."
 RN Comp. Biochem. Physiol. 93B:929-934(1989).
 RN [7]
 RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=98062367; PubMed=9398529;
 RA Moore S.A., Anderson B.F., Groom C.R., Haridas M., Baker E.N.;
 RT "Three-dimensional structure of ferric bovine lactoferrin at 2.8-A
 RL resolution."
 RL J. Mol. Biol. 274:222-236(1997).
 RN [8]
 RN STRUCTURE BY NMR OF 36-60.
 RX MEDLINE=98190007; PubMed=9521752;
 RA Hwang P.M., Zhou N., Shan X., Artowsmith C.H., Vogel H.J.;
 RT "Three-dimensional solution structure of lactoferricin B, an
 RL antimicrobial peptide derived from bovine lactoferrin."
 RL Biochemistry 37:4288-4298(1998).
 CC - FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE.
 CC - FUNCTION: LACTOFERRICIN B IS AN ANTIMICROBIAL PEPTIDE.
 CC - SUBUNIT: Monomer.
 CC - SUBCELLULAR LOCATION: Secreted.
 CC - DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC - SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage fee and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X57084; CAA40366.1; -
 DR EMBL; M63502; AAA30617.1; -
 DR EMBL; L08604; AAA30609.1; -
 DR EMBL; L19983; AAA21722.1; -
 DR EMBL; L19982; AAA21722.1; JOINED.
 DR EMBL; L19983; AAA21722.1; JOINED.
 DR EMBL; L19984; AAA21722.1; JOINED.
 DR EMBL; L19985; AAA21722.1; JOINED.
 DR EMBL; L19986; AAA21722.1; JOINED.
 DR EMBL; L19988; AAA21722.1; JOINED.
 DR EMBL; L19989; AAA21722.1; JOINED.
 DR EMBL; L19990; AAA21722.1; JOINED.
 DR EMBL; L19991; AAA21722.1; JOINED.
 DR EMBL; L19992; AAA21722.1; JOINED.
 DR EMBL; AB046664; BAB03470.1; -

[illegible]

DT 21-UTL-1986 (Rel. 01, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lactoferrin precursor (Lactoferrin) [Contains: Lactoferrin A;
 DE Lactoferrin B; Lactoferrin C].
 GN LTF OR LF.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=90384839; PubMed=2402455;
 RA Rey M.W., Moloshuk S.L., de Boer H.A., Pieper F.R.;
 RT "Complete nucleotide sequence of human mammary gland lactoferrin.";
 RL Nucleic Acids Res. 18:5288-5288(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Cho Y.Y.;
 RL Thesis (1994), Genetic Engineering Research Institute / Taejon, Korea.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC Connolly O.M.;
 RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Liang Q., Jimenez-Flores R., Richardson T.;
 RT "Molecular cloning and sequence analysis of human lactoferrin.";
 RL Submitted (DEC-1991) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RA Wei X., Han J., Rado T.A.;
 RT "Human neutrophil lactoferrin coding and 5' flanking region DNA
 sequences.";
 RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Cheng H., Chen X., Huan L.;
 RT "cDNA cloning and sequence analysis of human lactoferrin.";
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Prostate;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Mariani K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshylyuk S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muliyil S.J.,
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Hulton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.U.W., Marra W.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [8]
 RP SEQUENCE OF 3-711 FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=90326549; PubMed=2374734;
 RA Powell M.J., Ogden J.E.;
 RT "Nucleotide sequence of human lactoferrin cDNA.";
 RL Nucleic Acids Res. 18:4013-4013(1990).
 RN [9]
 RP SEQUENCE OF 20-711.
 RX MEDLINE=85076667; PubMed=6510420;
 RA Metz-Boutigue M.-H., Jolles J., Mazurier J., Schoentgen F.,
 RA Legrand D., Spik G., Montreuil J., Jolles P.;
 RT "Human lactoferrin: amino acid sequence and structural
 comparisons with other transferrins.";
 RL Eur. J. Biochem. 145:659-666(1984).
 RN [10]
 RP PRELIMINARY SEQUENCE OF 20-73, 134-171, 257-278 AND 347-530.
 RX MEDLINE=82046817; PubMed=6794640;
 RA Metz-Boutigue M.-H., Mazurier J., Jolles J., Spik G., Montreuil J.,
 RA Jolles P.;
 RT "The present state of the human lactoferrin sequence. Study and
 alignment of the cyanogen bromide fragments and characterization of
 N- and C-terminal domains.";
 RL Biochim. Biophys. Acta 670:243-254(1981).
 RN [11]
 RP SEQUENCE OF 609-711.
 RX MEDLINE=82262043; PubMed=7049727;
 RA Metz-Boutigue M.-H., Jolles J., Mazurier J., Spik G., Montreuil J.,
 RA Jolles P.;
 RT "An 88 amino acid long C-terminal sequence of human
 lactoferrin.";
 RL FEBS Lett. 142:107-110(1982).
 RN [12]
 RP SEQUENCE OF 436-711 FROM N.A.
 RX MEDLINE=88001031; PubMed=3477300;
 RA Rado T.A., Wei X., Benz E.J., Jr.;
 RT "Isolation of lactoferrin cDNA from a human myeloid library and
 expression of mRNA during normal and leukemic myelopoiesis.";
 RL Blood 70:989-993(1987).
 RN [13]
 RP SEQUENCE OF 237-711 FROM N.A.
 RA McCombie W.R., Wilson R., Chen E., Gibbs R., Zuo L., Johnson D.,
 RA Nhan M., Parnell L., Dedhia N., Ansari A., Mardis E., Schutz K.,
 RA Gnoj L., La Baetide M., Kaplan N., Greco T., Touchman J., Muzny D.,
 RA Chen C.N., Evans C., Fitzgerald M., See L.H., Tang M., Porcel B.M.,
 RA Dragen Y., Giacalone J., Pae A., Powell E., Solinsky K.A., Desilva U.,
 RA Diaz-Perez S., Zhou X., Yu Y., Watanabe M., Doggett N., Garcia D.,
 RA Sagripanti J.L.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [14]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND REVISIONS.
 RX MEDLINE=90064528; PubMed=2585506;
 RA Anderson B.F., Baker H.M., Norris G.E., Rice D.W., Baker E.N.;
 RT "Structure of human lactoferrin: crystallographic structure analysis
 and refinement at 2.8-A resolution.";
 RL J. Mol. Biol. 209:711-734(1989).
 RN [15]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RA Haridas M., Anderson B.F., Baker E.N.;
 RT "Structure of human diferric lactoferrin refined at 2.2-A
 resolution.";
 RL Acta Crystallogr. D 51:629-646(1995).
 RN [16]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 20-353.
 RX MEDLINE=97156796; PubMed=9003186;
 RA Nicholson H., Anderson B.F., Bland T., Shewry S.C., Tweedie J.W.,
 RA Baker E.N.;
 RT "Mutagenesis of the histidine ligand in human lactoferrin: iron
 binding properties and crystal structure of the histidine-
 253->methionine mutant.";
 RL Biochemistry 36:341-346(1997).
 RN [17]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=99190892; PubMed=10089347;
 RA Sun X.L., Baker H.M., Shewry S.C., Jameson G.B., Baker E.N.;
 RT "Structure of recombinant human lactoferrin expressed in Aspergillus
 awamori.";

RL Acta Crystallogr. D 55:403-407(1999).
 RN [18]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=99192677; PubMed=10089508;
 RA Jameson G.B., Anderson B.F., Norris G.E., Thomas D.H., Baker E.N.;
 RT "Structure of human apolactoferrin at 2.0-A resolution. Refinement
 RL and analysis of ligand-induced conformational change.";
 RN Acta Crystallogr. D 54:1319-1335(1998).
 [19]
 RP CHARACTERIZATION OF LACTOFERROXINS.
 RX MEDLINE=91166929; PubMed=13569293;
 RA Tani F., Iio K., Chiba H., Yoshikawa M.;
 RT "Isolation and characterization of opitoid antagonist peptides derived
 RL from human lactoferrin.";
 RN Agric. Biol. Chem. 54:1803-1810(1990).
 [20]
 RP VARIANTS THR-30 AND ARG-48.
 RX MEDLINE=99091914; PubMed=9873069;
 RA Klintworth G.K., Sommer J.R., Obrian G., Han L., Ahmed M.N.,
 RA Qumailiyeh M.B., Lin P.-Y., Basti S., Reddy M.K., Kanai A., Hotta Y.,
 RA Sugar J., Kumaramanchavel G., Munier P., Schorderet D.F.,
 RA El Materi L., Iwata F., Kaiser-Kupfer M., Nagata M., Nakayasu K.,
 RA Hejtmancik J.F., Teng C.T.;
 RT "Familial subepithelial corneal amyloidosis (gelatinous drop-like
 RL corneal dystrophy): exclusion of linkage to lactoferrin gene.";
 RN Mol. Vision 4:31-32(1998).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE.
 CC -1- FUNCTION: LACTOFERROKINS A, B AND C HAVE OPIOID ANTAGONIST
 CC ACTIVITY. LACTOFERROKIN A SHOWS PREFERENCE FOR MU-RECEPTORS, WHILE
 CC LACTOFERROKIN B AND C HAVE SOMEWHAT HIGHER DEGREES OF PREFERENCE
 CC FOR KAPPA-RECEPTORS THAN FOR MU-RECEPTORS.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X53961; CA37914.1; -;
 DR EMBL; U07643; AAB60324.1; -;
 DR EMBL; M93150; AAA36159.1; -;
 DR EMBL; M83202; AAA59511.1; -;
 DR EMBL; M83205; AAA58656.1; -;
 QY 1 PMSKCYQWQNRKXIGAPSTICIRTS 28
 Db 34 PEAIKCFQWQNRKXVGPVSCIKRDS 61
 Query Match 61.0%; Score 100; DB 1; Length 711;
 Best Local Similarity 57.1%; Pred. No. 4.2e-07;
 Matches 16; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RESULT 6
 TRFL MOUSE
 ID TRFL MOUSE STANDARD; PRT; 707 AA.
 AC P08071; P70690; Q61799; Q922P2;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lactotransferrin precursor (lactoferrin).
 GN LTF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RX MEDLINE=87280033; PubMed=3611056;
 RA Pentecost B.T., Teng C.T.;
 RT "Lactotransferrin is the major estrogen inducible protein of mouse
 RL uterine secretions.";
 RN J. Biol. Chem. 262:10134-10139(1987).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Uterus;
 RA Morishita K.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettlemen M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences.";
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [4]
 RP SEQUENCE OF 1-14 FROM N.A.
 RX MEDLINE=92042099; PubMed=1939212;
 RA Liu Y., Teng C.T.;
 RT "Characterization of estrogen-responsive mouse lactoferrin promoter.";
 RL J. Biol. Chem. 266:21880-21885(1991).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; J03298; AAA40525.1; -;
 DR EMBL; D88510; BAA13633.1; -;
 DR EMBL; BC006904; AAH06904.1; -;
 DR EMBL; M74778; AAA39427.1; -;
 DR HSSP; P02788; IC86.
 DR MGD; MGI:96837; Ltf.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_PRR_2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 1.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;

KM Signal. 1 19 BY SIMILARITY.
 FT SIGNAL. 20 707 LACTOTRANSFERRIN.
 FT CHAIN 20 357 1.
 FT REPEAT 358 707 2.
 FT DISULFID 27 63 BY SIMILARITY.
 FT DISULFID 37 54 BY SIMILARITY.
 FT DISULFID 133 216 BY SIMILARITY.
 FT DISULFID 175 191 BY SIMILARITY.
 FT DISULFID 188 199 BY SIMILARITY.
 FT DISULFID 249 263 BY SIMILARITY.
 FT DISULFID 366 398 BY SIMILARITY.
 FT DISULFID 376 389 BY SIMILARITY.
 FT DISULFID 423 702 BY SIMILARITY.
 FT DISULFID 443 665 BY SIMILARITY.
 FT DISULFID 475 550 BY SIMILARITY.
 FT DISULFID 499 693 BY SIMILARITY.
 FT DISULFID 509 523 BY SIMILARITY.
 FT DISULFID 520 533 BY SIMILARITY.
 FT DISULFID 591 605 BY SIMILARITY.
 FT DISULFID 643 648 BY SIMILARITY.
 FT METAL 78 78 IRON 1 (BY SIMILARITY).
 FT METAL 110 110 IRON 1 (BY SIMILARITY).
 FT METAL 210 210 IRON 1 (BY SIMILARITY).
 FT METAL 271 271 IRON 1 (BY SIMILARITY).
 FT METAL 413 413 IRON 2 (BY SIMILARITY).
 FT METAL 451 451 IRON 2 (BY SIMILARITY).
 FT METAL 544 544 IRON 2 (BY SIMILARITY).
 FT METAL 613 613 IRON 2 (BY SIMILARITY).
 FT BINDING 139 139 ANION (POTENTIAL).
 FT BINDING 481 481 ANION (POTENTIAL).
 FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 494 494 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1 2 MR -> IQG (IN REF. 1).
 FT CONFLICT 25 25 R -> Q (IN REF. 2).
 FT CONFLICT 82 82 M -> L (IN REF. 2).
 FT CONFLICT 359 359 S -> T (IN REF. 2).
 FT CONFLICT 382 382 A -> D (IN REF. 1).
 FT CONFLICT 449 449 E -> G (IN REF. 2).
 FT CONFLICT 629 629 L -> V (IN REF. 1).
 SQ SEQUENCE 707 AA; 77865 MW; F26AE0340A4C19A8 CRC64;

Query Match 51.2%; Score 84; DB 1; Length 707;
 Best Local Similarity 44.4%; Pred. No. 9.8e-05;
 Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 EWSKCYOMQRMRKLGAPSTICIRRTS 28
 DB 33 EEEKCLRMQEMRKVGGPPLSCVKKS 59

RESULT 7
 ID TRFL_HORSE STANDARD; PRT; 695 AA.
 AC 077811;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-SRP-2003 (Rel. 42, Last annotation update)
 DE Lactotransferrin precursor (lactoferrin) (fragment).
 GN LTF.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
 OX NCBI_TaxID=9796;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Paramasivam M., Srinivasan A., Yadav M.P., Singh T.P.;
 RT "cDNA sequence of mare lactoferrin";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RC TISSUE=Milk;
 RX MEDLINE=99296631; PubMed=1036507;

RA Sharma A.K., Paramasivam M., Srinivasan A., Yadav M.P., Singh T.P.;
 RT "three-dimensional structure of mare deferitic lactoferrin at 2.6-A
 resolution";
 RL J. Mol. Biol. 289:303-317(1999).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ010930; CA093407.1; -.
 CC PDB; 1BX; 02-DEC-98.
 CC PDB; 1B7U; 02-FEB-99.
 CC PDB; 1B7Z; 02-FEB-99.
 CC PDB; 1F9B; 10-FEB-01.
 CC PDB; 1IEB; 13-FEB-02.
 CC PDB; 1QJM; 14-JAN-00.
 CC InterPro; IPR001156; Transferrin.
 CC Pfam; PF00405; Transferrin; 2.
 CC PRINTS; PR00422; TRANSFERRIN.
 CC SMART; SM00094; TR_FER; 2.
 CC PROSITE; PS00205; TRANSFERRIN_1; 2.
 CC PROSITE; PS00206; TRANSFERRIN_2; 2.
 CC PROSITE; PS00207; TRANSFERRIN_3; 1.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.
 KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal; 3D-structure.
 FT SIGNAL 1 6
 FT CHAIN 7 695 LACTOTRANSFERRIN.
 FT REPEAT 7 350 1.
 FT REPEAT 351 695 2.
 FT DISULFID 15 51
 FT DISULFID 25 42
 FT DISULFID 121 204
 FT DISULFID 163 179
 FT DISULFID 176 187
 FT DISULFID 189 189
 FT DISULFID 237 251
 FT DISULFID 354 386
 FT DISULFID 377 377
 FT DISULFID 411 690
 FT DISULFID 431 653
 FT DISULFID 463 538
 FT DISULFID 487 681
 FT DISULFID 497 511
 FT DISULFID 508 521
 FT DISULFID 579 593
 FT DISULFID 631 636
 FT METAL 66 66
 FT METAL 98 98
 FT METAL 198 198
 FT METAL 259 259
 FT METAL 401 401
 FT METAL 439 439
 FT METAL 532 532
 FT METAL 601 601
 FT METAL 631 631
 FT BINDING 127 127
 FT BINDING 143 143
 FT CARBOHYD 287 287
 FT CARBOHYD 482 482
 FT STRAND 12 16
 FT HELIX 19 34
 IRON 1 (BY SIMILARITY).
 IRON 1 (BY SIMILARITY).
 IRON 1 (BY SIMILARITY).
 IRON 1 (BY SIMILARITY).
 IRON 2 (BY SIMILARITY).
 IRON 2 (BY SIMILARITY).
 IRON 2 (BY SIMILARITY).
 ANION (BY SIMILARITY).
 ANION (BY SIMILARITY).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).


```

FT      TURN      35      36
FT      STRAND     40      44
FT      HELIX      48      56
FT      TURN      57      58
FT      STRAND     62      65
FT      HELIX      67      74
FT      TURN      76      78
FT      STRAND     80      88
FT      STRAND     95      95
FT      STRAND     97      105
FT      TURN      106      107
FT      HELIX      112      114
FT      TURN      116      117
FT      STRAND     119      122
FT      TURN      125      126
FT      TURN      128      131
FT      HELIX      132      137
FT      HELIX      139      142
FT      TURN      147      148
FT      HELIX      151      158
FT      STRAND     161      163
FT      TURN      165      166
FT      TURN      169      171
FT      HELIX      173      175
FT      TURN      176      178
FT      TURN      183      187
FT      TURN      191      192
FT      TURN      194      195
FT      HELIX      197      206
FT      TURN      207      208
FT      STRAND     212      216
FT      HELIX      219      223
FT      HELIX      227      230
FT      TURN      231      232
FT      STRAND     233      237
FT      TURN      238      240
FT      STRAND     241      244
FT      HELIX      245      247
FT      TURN      248      250
FT      STRAND     254      257
FT      STRAND     260      264
FT      HELIX      270      284
FT      TURN      286      287
FT      TURN      294      295
FT      TURN      308      309
FT      STRAND     312      315
FT      HELIX      322      326
FT      HELIX      328      336
FT      TURN      337      338
FT      HELIX      341      349
FT      STRAND     351      356
FT      HELIX      358      370
FT      TURN      371      372
FT      STRAND     375      380
FT      HELIX      383      392
FT      TURN      393      393
FT      STRAND     397      400
FT      HELIX      402      410
FT      TURN      411      412
FT      STRAND     414      421
FT      HELIX      429      433
FT      STRAND     439      446
FT      HELIX      454      456
FT      TURN      458      459
FT      STRAND     462      464
FT      TURN      467      468
FT      TURN      470      473
FT      HELIX      474      484
FT      HELIX      489      491
FT      TURN      492      492
FT      STRAND     496      497
FT      TURN      499      500
FT      TURN      503      504

```

```

FT      TURN      506      507
FT      STRAND     508      508
FT      TURN      515      516
FT      TURN      520      521
FT      TURN      525      526
FT      TURN      528      529
FT      HELIX      531      540
FT      TURN      541      542
FT      STRAND     546      550
FT      HELIX      551      555
FT      TURN      556      557
FT      TURN      559      560
FT      HELIX      565      568
FT      TURN      569      569
FT      TURN      572      573
FT      STRAND     575      578
FT      TURN      580      581
FT      STRAND     584      586
FT      HELIX      587      592
FT      STRAND     596      598
FT      STRAND     602      605
FT      TURN      607      609
FT      HELIX      610      624
FT      TURN      626      627
FT      TURN      629      630
FT      HELIX      631      634
FT      TURN      637      638
FT      TURN      641      642
FT      TURN      649      650
FT      STRAND     651      655
FT      TURN      658      659
FT      HELIX      663      667
FT      HELIX      669      679
FT      TURN      680      681

```

Query Match Best Local Similarity 50.6%; Score 83; DB 1; Length 695;
Matches 14; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 2 EWSKCYQWORMRKAGPSITCIRRTSA 29
Db 21 EAKCAKFGQNMKXVGPVSCIRRTSS 48

```

RESULT 8
TRFL_PIG STANDARD; PRT; 704 AA.
AC P14632; Q29557;
DT 01-APR-1990 (Rel. 14, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactoferrin precursor (Lactoferrin).
LTF.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92379101; PubMed=1511016;
RA Lyndon J.P., O'Malley B.R., Saucedo O., Lee T., Heaton D.R.,
RA Connely O.M.;
RT "Nucleotide and primary amino acid sequence of porcine lactoferrin.";
RL Biochim. Biophys. Acta 1132:97-99(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92367939; PubMed=1503259;
RA Alexander L.J., Levine W.B., Teng C.T., Beattie C.W.;
RT "Cloning and sequencing of the porcine lactoferrin cDNA.";
RL Anim. Genet. 23:251-256(1992).
RN [3]
RP SEQUENCE OF 20-49.
RX MEDLINE=90105538; PubMed=2605266;

```


DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 15-SEP-2003 (Rel. 42, last annotation update)
 DE Putative mannosyl-3-phosphoglycerate phosphatase (EC 3.1.3.70) (MGP)
 GN YNDP OR STM1986 OR STY2193 OR T0892.
 OS Salmomella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmomella.
 OX NCBI_TaxId=602, 601;
 [1]
 RN SEQUENCE FROM N.A.
 RP SPECIES=S.typhimurium; STRAIN=L72 / SSGC1412 / ATCC 700720;
 RC MEDLINE=21534948; PubMed=11677609;
 RA McCallum M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan B., Sun H., Florza L., Miller W., Stoneking T., Nian M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmomella enterica serovar Typhimurium
 RT L72."
 RL Nature 413:852-856(2001).
 [2]
 RN SEQUENCE FROM N.A.
 RP SPECIES=S.typhi; STRAIN=CT18;
 RC MEDLINE=21534947; PubMed=11677608;
 RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Dougan K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltham T., Hamlin N., Haque A., Hien T.T., Holtz S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmomella
 RT enterica serovar Typhi CT18."
 RL Nature 413:848-852(2001).
 [3]
 RN SEQUENCE FROM N.A.
 RP SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;
 RC MEDLINE=22531367; PubMed=12644504;
 RA Deng W., Lion S.-R., Plumet G. III, Mayhew G.F., Rose D.J.,
 RA Burdand V., Kodoyanni V., Schwartz D.C., Blatner F.R.;
 RT "Comparative genomics of Salmomella enterica serovar Typhi strains Ty2
 RT and CT18."
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -1- CATALYTIC ACTIVITY: 2(alpha-D-mannosyl)-3-phosphoglycerate + H(2)O
 CC = 2(alpha-D-mannosyl)-D-glycerate + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HAD SUPERFAMILY. MGP FAMILY.
 CC "This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB008788; AL20896.1; -
 CC EMBL; AL627272; CAD05733.1; -
 CC EMBL; AB016837; AAC068570.1; -
 CC SrvGene; SGR7777; Yedp.
 CC HAMAP; MF_00617; -1.
 CC InterPro; IPR006379; HAD_SF_IIB.
 CC InterPro; IPR006381; HAD_SF_Yedp.
 CC InterPro; IPR005834; Hydrolase.
 CC Pfam; PF00702; Hydrolase; 1.
 CC TIGRFAMs; TIGR01486; HAD-SF-IIB-Yedp; 1.
 CC TIGRFAMs; TIGR01484; HAD-SF-IIB; 1.
 CC Hypothetical protein; Hydrolase; Complete proteome.
 CC SQUENCE 271 AA; 30925 MW; 46866C2B36B7008E CRC64;
 S0

Query Match 35.4%; Score 58; DB 1; Length 271;

Best Local Similarity 35.7%; Pred. No. 0.25;
 Matches 10; Conservative. 4; Mismatches 14; Indels 0; Gaps 0;
 QY 2 EWSKQWQRMKXGAPSTICIRTS 29
 DB 25 EWQAPAPWLRHSGVPVILCSKTA 52
 RESULT 13
 TRFM_HUMAN STANDARD; PRT; 738 AA.
 AC P08562; Q990E2;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Melanotransferrin precursor (Melanoma-associated antigen p97) (CD228
 DE antigen).
 GN MF12 OR MAP97.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RN SEQUENCE FROM N.A. (ISOFORM 1).
 RP TISSUE=Melanoma;
 RC MEDLINE=86149285; PubMed=2419904;
 RA Rose T.M., Ploewman G.D., Teplow D.B., Dreyer W.J., Hellstroem K.E.,
 RA Brown J.P.;
 RT "Primary structure of the human melanoma-associated antigen p97
 RT (melanotransferrin) deduced from the mRNA sequence."
 RL Proc. Natl. Acad. Sci. U.S.A. 83:1261-1265(1986).
 [2]
 RN SEQUENCE FROM N.A. (ISOFORM 2).
 RP TISSUE=Skin, and Uterus;
 RC MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Heath N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heist F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Staretz M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.U., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millhys S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gumaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Feeney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [3]
 RN GPI-ANCHOR.
 RP MEDLINE=94132080; PubMed=8300636;
 RA Food W.R., Rothemburger S., Gabathuler R., Haidl I.D., Reid G.,
 RA Jeffries W.A.;
 RT "Transposon and expression in human melanomas of a transferrin-like
 RT glycosylphosphatidylinositol-anchored protein."
 RL J. Biol. Chem. 269:3034-3040(1994).
 [4]
 RN FUNCTION.
 RP MEDLINE=96016189; PubMed=7556058;
 RA Kennard M.L., Richardson D.R., Gabathuler R., Ponka P.,
 RA Jeffries W.A.;
 RT "A novel iron uptake mechanism mediated by GPI-anchored human p97."
 RL EMBO J. 14:4178-4186(1995).
 [5]
 RN IRON-BINDING.
 RP MEDLINE=92183668; PubMed=1544447;
 RX

RA Baker E.N., Baker H.M., Smith C.A., Stebbins M.R., Kahn M.,
 RA Hellstrom K.E., Hellstrom I.,
 RA "Human melanotransferrin (p97) has only one functional iron-binding
 RL site.";
 RL FEBS Lett. 298:215-218 (1992).
 RN [6]
 RX 3D-STRUCTURE MODELING.
 RX MEDLINE=92339524; PubMed=1633859;
 RA Garret R.C., Jochi H.;
 RA "A molecular model for the tumour-associated antigen, p97, suggests a
 RT Zn-binding function";
 RL FEBS Lett. 305:55-61(1992).
 CC -1- FUNCTION: INVOLVED IN IRON CELLULAR UPTAKE. SEEMS TO BE
 CC INTERNALIZED AND THEN RECYCLED BACK TO THE CELL MEMBRANE. BINDS A
 CC SINGLE ATOM OF IRON PER SUBUNIT. COULD ALSO BIND ZINC.
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=P08582-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=P08582-2; Sequence=VSP_006557; VSP_006558;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: FOUND PREDOMINANTLY IN HUMAN MELANOMAS AND IN
 CC CERTAIN FETAL TISSUES; ALSO FOUND IN LIVER, EPITHELIUM, UMBILICAL
 CC CHORD, PLACENTA AND SWEAT GLAND DUCTS.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M12154; AAA5992.1; -;
 DR EMBL; A00127; CAA00012.1; -;
 DR EMBL; BC001875; AAH01875.1; -;
 DR EMBL; BC002623; AAH02623.1; -;
 DR EMBL; BC007550; AAH07550.1; -;
 DR PIR; A23814; TTHOM.
 DR HSSP; P19134; ITPD.
 DR GeneW; HGNC:7037; MF12.
 DR MIM; 155750; -;
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0008222; F:tumor antigen; TAS.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SM00094; TR_FRR; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 DR KEGG; K00000; Transferrin; Glycoprotein; Metal-binding; Transmembrane;
 KW Repeat; Signal; GPI-anchor; Membrane; Zinc; Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 ? MELANOTRANSFERRIN.
 FT PROPEP 20 738 REMOVED IN MATURE FORM.
 FT REPEAT 20 361 1.
 FT DISULFID 362 713 2.
 FT DISULFID 26 63 BY SIMILARITY.
 FT DISULFID 36 54 BY SIMILARITY.
 FT DISULFID 130 216 BY SIMILARITY.
 FT DISULFID 172 189 BY SIMILARITY.
 FT DISULFID 186 199 BY SIMILARITY.
 FT DISULFID 257 271 BY SIMILARITY.
 FT METAL 78 78 IRON 1 (BY SIMILARITY).
 FT METAL 107 107 IRON 1 (BY SIMILARITY).
 FT METAL 210 210 IRON 1 (BY SIMILARITY).
 FT METAL 279 279 IRON 1 (BY SIMILARITY).
 FT METAL 421 421 IRON 2 (BY SIMILARITY).
 CC

FT METAL 451 451 IRON 2 (BY SIMILARITY).
 FT METAL 556 556 IRON 2 (BY SIMILARITY).
 FT METAL 625 625 IRON 2 (BY SIMILARITY).
 FT BINDING 136 136 ANION (POTENTIAL).
 FT TRANSMEM 714 738 ANCHOR.
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 238 302 GKLPSGQALLSDPEFLCRGSRADVTETMRQCHARYPA
 FT HAVVRAPIIDGGLITRLNLEGQRL -> ESRPRQRTWSE
 FT EEEGCPAHEERARTTRSSAGQAMKAPVHRPQDESCKGEF
 FT GRARSRDLG (in isoform 2).
 FT /FTID=VSP_006557.
 FT /FTID=VSP_006558.
 FT /FTID=VSP_006559.
 FT VARSPLIC 303 738 Missing (in isoform 2).
 FT SEQUENCE 738 AA; 80241 MW; 6E6086B94D7B955 CRC64;
 SQ
 Query Match 35.1%; Score 57.5; DB 1; Length 738;
 Best Local Similarity 43.3%; Pred. No. 0.85;
 Matches 13; Conservative 3; Mismatches 13; Indels 1; Gaps 1;
 QY 1 PEMSKCYQWRMRKLG-APSTICRTRTSA 29
 DB 31 PEQHKCGNNSAFREAGIQPSLLCVRGTS 60
 RESULT 14
 ID PEMT_MOUSE STANDARD; PRT; 198 AA.
 AC Q61907;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phosphatidyldethanolamine N-methyltransferase (EC 2.1.1.17) (PEMT)
 DE (PEMT2).
 OS Mus musculus (Mouse).
 GN PEMT OR PEMT2.
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=129/J; TISSUE=Liver;
 RX MEDLINE=97133127; PubMed=8978486;
 RA Walkey C.J., Cui Z., Agellon L.B., Vance D.B.;
 RT "Characterization of the murine phosphatidyldethanolamine
 RT N-methyltransferase-2 gene.";
 RL J. Lipid Res. 37:2341-2350(1996).
 RN (2)
 RP CHARACTERIZATION.
 RX MEDLINE=98058738; PubMed=9371769;
 RA Walkey C.J., Donohue L.R., Bronson R., Agellon L.B., Vance D.E.;
 RT "Disruption of the murine gene encoding phosphatidyldethanolamine
 RT N-methyltransferase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:12880-12885(1997).
 RN (3)
 RP CHARACTERIZATION.
 RX MEDLINE=98438461; PubMed=9765216;
 RA Walkey C.J., Yu L., Agellon L.B., Vance D.E.;
 RT "Biochemical and evolutionary significance of phospholipid
 RT methylation.";
 RL J. Biol. Chem. 273:27043-27046(1998).
 CC -1- FUNCTION: CATALYZE THREE SEQUENTIAL METHYLATION OF
 CC PHOSPHATIDYLETHANOLAMINE (PE) BY ADOMET, THUS PRODUCING
 CC PHOSPHATIDYLCHOLINE (PC).
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine +
 CC phosphatidyl-N-methylethanolamine = S-adenosyl-L-homocysteine +
 CC phosphatidyl-N-methylethanolamine.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL). FOUND
 CC IN MITOCHONDRIA AND IN ENDOPLASMIC RETICULUM.
 CC -1- TISSUE SPECIFICITY: LIVER.
 CC -1- SIMILARITY: BELONGS TO THE PEMT / PEM2 METHYLTRANSFERASE FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U25051; AAA67686.1; -;
 DR EMBL; U25046; AAA67686.1; JOINED.
 DR EMBL; U25047; AAA67686.1; JOINED.
 DR EMBL; U25048; AAA67686.1; JOINED.
 DR EMBL; U25049; AAA67686.1; JOINED.
 DR EMBL; U25050; AAA67686.1; JOINED.
 DR MGD; MGI:104535; Pent.
 DR Pfam; PF04191; PENT.1.
 DR Phospholipid biosynthesis; Transferase; Methyltransferase;
 DR Transmembrane; Mitochondrion; Endoplasmic reticulum.
 FT INIT MET 0 0
 FT TRANSMEM 12 32 POTENTIAL.
 FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 157 177 POTENTIAL.
 SQ SEQUENCE 198 AA; 22385 MW; 5828975A55173D72 CRC64;

Query Match 31.7%; Score 52; DB 1; Length 198;
 Best Local Similarity 36.0%; Pred. No. 1.4;
 Matches 9; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

QY 3 WSKCYQWRMRKRL---GAPSTTC 23
 Db 28 MNVAVRWGRTRKLSRAFGSPHLAC 52

RESULT 15
 MGP_ECOL6 STANDARD; PRT; 271 AA.
 AC P59286;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative mannosyl-3-phosphoglycerate phosphatase (EC 3.1.3.70) (MGP).
 GN YEDP OR C2373.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OK NCBI_TaxID=217992;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackert J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Domeneberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 RL -1- CATALYTIC ACTIVITY: 2(alpha-D-mannosyl)-3-phosphoglycerate + H(2)O
 CC = 2(alpha-D-mannosyl)-D-glycerate + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HAD SUPERFAMILY. MGP FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AE016762; AAN80832.1; -;
 DR HAMAP; MF_00617; -; 1.

DR Pfam; PF00702; Hydrolase; 1.
 DR TIGRPFAM; TIGR01486; HAD-SF-11B-YedP; 1.
 DR TIGRPFAM; TIGR01484; HAD-SF-11B; 1.
 KW Hypothetical protein; Hydrolase; Complete proteome.
 SQ SEQUENCE 271 AA; 30515 MW; E6EB78E391470B2 CRC64;

Query Match 31.7%; Score 52; DB 1; Length 271;
 Best Local Similarity 32.1%; Pred. No. 2;
 Matches 9; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 2 EWSKCYQWRMRKLGAPSTTCIRRTSA 29
 Db 25 DWQAPAPWLSRLHEANIPVILCSKRTSA 52

Search completed: July 30, 2003, 16:24:31
 Job time : 10.494 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 18.4702 seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-6
Perfect score: 164
Sequence: 1 PEMSKCYQWQRMRKLGAPSTICIRRTSA 29
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	99.4	708	2 JC2323	lactoferrin - goat
2	122	74.4	708	1 TFBOL	lactotransferrin p
3	100	61.0	711	1 TFMUL	lactotransferrin p
4	91	55.5	33	2 SS2107	lactoferrin - Snee
5	84	51.2	707	1 A28438	lactoferrin precursor
6	66	40.2	703	2 A45543	carbonic anhydrase
7	58.5	35.7	704	2 I47228	hypothetical prote
8	58	35.4	271	2 E85813	hypothetical prote
9	58	35.4	271	2 E90965	hypothetical prote
10	58	35.4	271	2 G64959	hypothetical prote
11	58	35.4	271	2 AB0754	conserved hypochet
12	57.5	35.1	738	1 TFMUL	melanotransferrin
13	51	31.1	199	2 A47353	phosphatidylethano
14	51	31.1	489	2 C70655	probable monooxyge
15	50	30.5	706	2 S33761	transferrin precur
16	49	29.9	405	2 AB1461	B. subtilis YabB p
17	49	29.9	408	2 AC1098	B. subtilis YabB p
18	49	29.9	507	2 G87466	tryptophan halogen
19	49	29.9	1721	2 T21214	hypothetical prote
20	48.5	29.6	711	2 AH0033	hypothetical prote
21	48	29.3	430	2 D86573	histidyl tRNA synt
22	48	29.3	430	2 H72052	histidyl tRNA synt
23	47.5	29.3	423	2 F86313	histidine-tRNA lig
24	47	28.7	98	2 S39410	F2H5.8 protein -
25	47	28.7	528	2 T21834	nosr protein - Par
26	47	28.7	785	2 T48252	ecceiferrum3 (CEB3)
27	47	28.7	822	2 AB2507	hypothetical prote
28	47	28.7	2671	2 A49873	inositol 1,4,5-tri
29	46.5	28.4	121	2 AH3147	hypothetical prote

30	46.5	28.4	694	1 TFRBP	transferrin precur
31	46	28.0	413	2 AH0780	probable oxidoredu
32	46	28.0	469	2 T26463	hypothetical prote
33	46	28.0	488	2 S36570	E2 protein - human
34	46	28.0	622	2 S61692	probable membrane
35	46	28.0	647	2 E82126	DNA topoisomerase
36	46	28.0	1082	2 T50650	AP-3 complex beta
37	45.5	27.7	384	2 E69050	GlcNAc-phosphatidy
38	45.5	27.7	550	2 T47158	hypothetical prote
39	45.5	27.7	620	2 T27008	hypothetical prote
40	45.5	27.7	698	1 TFMUL	transferrin precur
41	45	27.4	200	1 A42725	nitrite hydratase
42	45	27.4	284	2 S58650	hypothetical prote
43	45	27.4	310	2 C69218	hypothetical prote
44	45	27.4	368	2 I55961	MHC class I histoc
45	45	27.4	395	2 T31578	hypothetical prote

ALIGNMENTS

RESULT 1
JC2323
lactoferrin - goat
C/Species: Capra aegagrus hircus (domestic goat)
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 07-May-1999
C/Accession: JC2323
R/Le Provost, F.; Nocart, M.; Guerin, G.; Martin, P.
Biochem. Biophys. Res. Commun. 203, 1324-1332, 1994
A/Title: Characterization of the goat lactoferrin cDNA: assignment of the relevant locus
A/Reference number: JC2323; MUID:94380047; PMID:8093048
A/Accession: JC2323
A/Molecule type: mRNA
A/Residues: 1-708 <LEP>
C/Superfamily: transferrin; transferrin repeat homology
C/Keywords: duplication; glycoprotein
F/359-696/Domain: transferrin repeat homology <TRH2>
F/252,300,387,495,564/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 99.4%; Score 163; DB 2; Length 708;
Best Local Similarity 96.6%; Pred. No. 7.3e-16;
Matches 28; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PEMSKCYQWQRMRKLGAPSTICIRRTSA 29
Db 33 PEMSKCYQWQRMRKLGAPSTICIRRTSA 61

RESULT 2
TFBOL
lactotransferrin precursor - bovine
N/Alternate names: lactoferrin
C/Species: Bos primigenius taurus (cattle)
C/Date: 31-Mar-1992 #sequence_revision 21-Nov-1997 #text_change 11-May-2000
C/Accession: I45919; S14674; S14110; S18517; U05055; S13097; S18518; S13881; PLO148; S22
R/Tsang, T.C.; Burns, D.K.; Wang, F.; Pan, Y.
FASEB J. 6, 233, 1991
A/Title: Cloning of a 80-kD advanced glycosylation end product (AGE) binding protein from
A/Reference number: I45919
A/Accession: I45919
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-708 <RSA>
A/Cross-references: GB:I08604; NID:G163269; PIDN:AAA30609.1; PID:G163270
R/Pierre, A.
Submitted to the EMBL Data Library, November 1990
A/Reference number: S14674
A/Accession: S14674
A/Molecule type: mRNA
A/Residues: 1-144 'V', 146-163 'P', 166-339 'A', 341-438 'Y', 440-513 'R', 515-708 <PI1>
A/Cross-references: EMBL:X57084; NID:G505; PIDN:CAA40366.1; PID:G506
R/Pierce, A.; Colavizza, D.; Benaisa, M.; Maes, P.; Tartar, A.; Montreuil, J.; Spik, G.
Eur. J. Biochem. 196, 177-184, 1991

A/Title: Molecular cloning and sequence analysis of bovine lactoferrin.
 A/Reference number: S14110; MUID:91160550; PMID:2001696
 A/Accession: S14110
 A/Molecule type: mRNA
 A/Residues: 3-144, 'V', 146-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PI2>
 A/Cross-references: EMBL:X57084
 A/Accession: S18517
 A/Molecule type: protein
 A/Residues: 20-35;82-114;148-163, 'PP', 166-178, 'V', 'P', 183-190;205-212;230-239;304-339;55
 R/Goodman, R.E.; Schanbacher, F.L.
 Biochem. Biophys. Res. Commun. 180, 75-84, 1991
 A/Title: Bovine lactoferrin mRNA: sequence, analysis, and expression in the mammary gland
 A/Reference number: J10595; MUID:92028886; PMID:1718281
 A/Accession: J10595
 A/Molecule type: mRNA
 A/Residues: 1-65, 'PG', 68-296, 'S', 298-339, 'A', 341-708 <GOO>
 A/Cross-references: GB:M63502
 A/Note: the authors translated the codon CCG for residue 66 as Arg and TCT for residue 2
 R/Mead, P.E.; Tweedle, J.W.
 Nucleic Acids Res. 18, 7167, 1990
 A/Title: cDNA and protein sequence of bovine lactoferrin.
 A/Reference number: S13097; MUID:91088328; PMID:2263492
 A/Accession: S13097
 A/Molecule type: mRNA
 A/Residues: 28-33, 'DS', 36-38, 'P', 40-708 <MEA>
 A/Cross-references: EMBL:X54801
 A/Accession: S18518
 A/Molecule type: protein
 A/Residues: 20-47;59-66;132-139;256-277;278,305-332;343-351;361-363;586,587-589;598-619
 R/Mead, P.E.
 Submitted to the EMBL Data Library, October 1990
 A/Reference number: S13881
 A/Accession: S13881
 A/Molecule type: mRNA
 A/Residues: 28-38, 'P', 40-86, 'C', 88-708 <ME3>
 A/Cross-references: EMBL:X54801
 R/Reiman, J.J.; Hegarty, H.M.; Hurley, M.L.
 Comp. Biochem. Physiol. B 93, 929-934, 1989
 A/Title: Purification and characterization of bovine lactoferrin from secretions of the
 A/Reference number: P10148; MUID:90031466; PMID:2805645
 A/Accession: P10148
 A/Molecule type: protein
 A/Residues: 20-27, 'X', 29-37, 'X', 39-54, 'X', 56-59 <REJ>
 R/Bellamy, W.; Takase, M.; Yamauchi, K.; Wakabayashi, H.; Kawase, K.; Tomita, M.
 Biochem. Biophys. Acta 1121, 130-136, 1992
 A/Title: Identification of the bactericidal domain of lactoferrin.
 A/Reference number: S21756; MUID:92287941; PMID:1599934
 A/Accession: S21756
 A/Molecule type: protein
 A/Residues: 36-60 <EBL>
 R/Shimazaki, K.; Tanaka, T.; Kon, H.; Oota, K.; Kawaguchi, A.; Maki, Y.; Sato, T.
 J. Dairy Sci. 76, 946-955, 1993
 A/Title: Separation and characterization of the C-terminal half molecule of bovine lacto
 A/Reference number: A56659; MUID:93253156; PMID:8486845
 A/Accession: A56659
 A/Molecule type: protein
 A/Residues: 20-25;302-308;359-366, 'X', 368-376, 'X', 378 <SHI>
 C/Superfamily: transferrin; transferrin repeat homology
 C/Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
 F/1-19/Domin: signal sequence #status predicted <SIG>
 F/20-708/Product: lactoferrin #status predicted <MAT>
 F/20-35/Domin: transferrin repeat homology <TRH1>
 F/36-60/Region: antimicrobial
 F/359-596/Domin: transferrin repeat homology <TRH2>
 F/28-64,134-217,176-192,179-200,189-202,250-264,367-399,377-390,424-703,444-666,476-551,
 F/38-55/Duplicate bonds: #status predicted
 F/7,11,21,272/Binding site: iron (Asp, Tyr, Tyr, His) #status experimental
 F/140/Binding site: carbonate (Arg) #status experimental
 F/222,300,387,495,564/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F/414,452,545,614/Binding site: iron (Asp, Tyr, Tyr, His) #status experimental
 F/482/Binding site: carbonate (Arg) #status experimental

Query Match

74.4%; Score 122; DB 1; Length 708;

Best Local Similarity 72.4%; Pred. No. 7,4e-10;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 PEMSKCYQWQRRKRGAPSTICIRRTSA 29
 Db 33 PEMFKCRWQWRKRGAPSTICVRPA 61

RESULT 3
 TFHUL
 lactoferrin precursor [validated] - human

N/Alternate names: lactoferrin

C/Species: Homo sapiens (man)

C/Date: 31-Mar-1992 #sequence revision 21-Nov-1997 #text change 08-Dec-2000

C/Accession: G01394; S11228; A45401; S10324; S15853; S20841; S07160; A61169; A31000; S74

R/Cho, Y.
 submitted to the EMBL Data Library, March 1994

A/Reference number: G06820

A/Accession: G01394

A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: mRNA

A/Residues: 1-711 <CHO>

A/Cross-references: EMBL:U07643; NID:9467236; PIDN:AA60324.1; PID:9467237

R/Rey, M.W.; Moloshuk, S.L.; deBoer, H.A.; Pieper, F.R.
 Nucleic Acids Res. 18, 5288, 1990

A/Title: Complete nucleotide sequence of human mammary gland lactoferrin.

A/Reference number: S11228; MUID:90384839; PMID:2402455

A/Accession: S11228

A/Molecule type: mRNA

A/Residues: 1-148, 'T', 150-422, 'C', 424-711 <REV>

A/Cross-references: EMBL:X53961; NID:934415; PIDN:CA37914.1; PID:934416

R/Teng, C.T.; Liu, Y.; Yang, N.; Walmer, D.; Panella, T.
 Mol. Endocrinol. 6, 1969-1981, 1992

A/Title: Differential molecular mechanism of the estrogen action that regulates lactofer

A/Reference number: A45401; MUID:93125571; PMID:1480183

A/Accession: A45401

A/Molecule type: DNA

A/Residues: 1-15 <TEN>

A/Cross-references: GB:S52659; NID:9263311; PIDN:AA624877.1; PID:9263312

A/Experimental source: placenta

A/Note: sequence extracted from NCBI backbone (NCBIP:122202)

R/Powell, M.J.; Ogden, J.E.
 Nucleic Acids Res. 18, 4013, 1990

A/Title: Nucleotide sequence of human lactoferrin cDNA.

A/Reference number: S10324; MUID:90326549; PMID:2374734

A/Accession: S10324

A/Molecule type: mRNA

A/Residues: 3-711 <POW>

A/Cross-references: EMBL:X52941; NID:934411; PIDN:CA37116.1; PID:934412

R/Stowell, K.M.; Rado, T.A.; Funk, W.D.; Tweedle, J.W.
 Biochem. J. 276, 349-355, 1991

A/Title: Expression of cloned human lactoferrin in baby-hamster kidney cells.

A/Reference number: S15853; MUID:91264786; PMID:2049066

A/Accession: S15853

A/Molecule type: mRNA

A/Status: nucleic acid sequence not shown; not compared with conceptual translation

A/Accession: S15853

A/Molecule type: mRNA

A/Residues: 20-31 <ST1>

A/Accession: S20841

A/Molecule type: protein

A/Residues: 20-28, 'X', 30-31 <ST2>

R/Rado, T.A.; Wei, X.; Benz Jr., E.J.
 Blood 70, 989-993, 1987

A/Title: Isolation of lactoferrin cDNA from a human myeloid library and expression of MR

A/Reference number: S07160; MUID:88001031; PMID:2477300

A/Accession: S07160

A/Molecule type: mRNA

A/Residues: 436-487, 'A', 489-711 <RAD>

A/Cross-references: EMBL:M18642; NID:9186815; PIDN:AA66665.1; PID:9386855

R/Panella, T.J.; Liu, Y.; Huang, A.T.; Teng, C.T.
 Cancer Res. 51, 3037-3043, 1991

A/Title: Polymorphism and altered methylation of the lactoferrin gene in normal leukocy

A/Reference number: A61169; MUID:91235214; PMID:1674448

A/Accession: A61169

A>Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 3-701, 'SMKPVN' <PAN>
 A:Experimental source: normal breast tissue
 R:Metz-Boutique, M.H.; Jolles, J.; Mazurier, J.; Schoengen, F.; Legrand, D.; Spik, G.;
 Eur. J. Biochem. 145, 659-666, 1984
 A>Title: Human lactoferrin: amino acid sequence and structural comparisons with oth
 A:Reference number: A31000; PMID:85076667; PMID:6510420
 A:Accession: A31000
 A:Molecule type: protein
 A:Residues: 20-140, 142-169, 171-203, 'L', 205, 'K', 207-208, 'K', 210-385, 'Q', 387-391, 'W', 393-4
 A>Note: this is the final paper in a series
 R:Houen, G.; Hoegdall, B.V.; Barkholt, V.; Nørskov, L.
 Eur. J. Biochem. 241, 303-308, 1996
 A>Title: Lactoferrin: similarity to diamine oxidase and purification by aminoethyl affini
 A:Reference number: S74119; PMID:97054624; PMID:8898921
 A:Accession: S74119
 A:Molecule type: protein
 A:Residues: 'G', 23-24, 'R', 26-27, 'XX', 30-32 <HOU>
 A:Experimental source: neutrophil granulocytes
 C:Genetics:
 A:Gene: GDB:LTP
 A:Cross-references: GDB:119368; OMIM:150210
 A:Map position: 3q21-3q23
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keyword: duplication; glycoprotein; iron binding; milk
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-71/Product: lactotransferrin #status experimental <MAT>
 F:21-356/Domain: transferrin repeat homology <TRH1>
 F:360-699/Domain: transferrin repeat homology <TRH2>
 F:59-65,35-58,135-218,177-193,190-201,251-265,503-697,595-609/disulfide bonds: #status e
 F:157,498/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:368-400,378-391,425-706,447-669,479-554,513-527,524-537,647-652/Disulfide bonds: #stat

Query Match 61.0%; Score 100; DB 1; Length 711;
 Best Local Similarity 57.1%; Pred. No. 1,2e-06;
 Matches 16; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 PEWSKCYQWRNRKLGAPSTICIRRTS 28
 DB 34 PEATKCFQWRNRKRGPPVSCIKRDS 61

RESULT 4
 S52107
 lactoferrin - sheep (fragment)
 C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
 C>Date: 14-Jul-1995 #sequence_revistion 05-Dec-1997 #text_change 05-Dec-1997
 C:Accession: S52107
 R:Qian, Z.Y.; Jolles, P.; Migliore-Samour, D.; Fiat, A.M.
 Biochim. Biophys. Acta 1243, 25-32, 1995
 A>Title: Isolation and characterization of sheep lactoferrin, an inhibitor of platelet a
 A:Reference number: S52107; PMID:95127729; PMID:7827104
 A:Accession: S52107
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-33 <QIA>
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication

Query Match 55.5%; Score 91; DB 2; Length 33;
 Best Local Similarity 78.9%; Pred. No. 1.4e-06;
 Matches 15; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 PEWSKCYQWRNRKLGAP 19
 DB 14 PEGSKCYQWRNRKLGAP 32

RESULT 5
 A28438
 lactoferrin precursor - mouse
 N:Alternate names: lactotransferrin

C:Species: Mus musculus (house mouse)
 C>Date: 10-Sep-1999 #sequence_revistion 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A28438; A41205
 R:Penecost, B.T.; Teng, C.T.
 J. Biol. Chem. 262, 10134-10139, 1987
 A>Title: Lactotransferrin is the major estrogen inducible protein of mouse uterine secre
 A:Reference number: A92596; PMID:8780033; PMID:3611056
 A:Accession: A28438
 A:Molecule type: mRNA
 A:Residues: 3-707 <PEN>
 A:Cross-references: EMBL:J03298
 R:Lin, Y.; Teng, C.T.
 J. Biol. Chem. 266, 21880-21885, 1991
 A>Title: Characterization of estrogen-responsive mouse lactoferrin promoter.
 A:Reference number: A41205; PMID:92042099; PMID:1939212
 A:Accession: A41205
 A:Molecule type: DNA
 A:Residues: 1-15 <LNU>
 A:Cross-references: GB:M74778
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-707/Product: lactotransferrin #status predicted <MAT>
 F:358-695/Domain: transferrin repeat homology <TRH2>
 F:494/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 51.2%; Score 84; DB 1; Length 707;
 Best Local Similarity 44.4%; Pred. No. 0.00027;
 Matches 12; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 2 EWSKCYQWRNRKLGAPSTICIRRTS 28
 DB 33 EEEKLRQWRNRKVGPPVSCVKKSS 59

RESULT 6
 A45543
 lactoferrin precursor - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 17-Feb-1994 #sequence_revistion 17-Feb-1994 #text_change 04-Mar-2000
 C:Accession: A45543; S24173
 R:Alexander, L.J.; Levine, W.B.; Teng, C.T.; Beattie, C.W.
 Anim. Genet. 23, 251-256, 1992
 A>Title: Cloning and sequencing of the porcine lactoferrin cDNA.
 A:Reference number: A45543; PMID:92367939; PMID:1503259
 A:Accession: A45543
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-703 <ALB>
 A:Experimental source: mammary gland
 A>Note: sequence extracted from NCBI backbone (NCBI:111151, NCBI:111153)
 R:Hydon, U.P.; O'Malley, B.R.; Saucedo, O.; Lee, T.; Headon, D.R.; Connely, O.M.
 Biochim. Biophys. Acta 1132, 97-99, 1992
 A>Title: Nucleotide and primary amino acid sequence of porcine lactoferrin.
 A:Reference number: S24173; PMID:92379101; PMID:1511016
 A:Accession: S24173
 A:Molecule type: mRNA
 A:Residues: 1-11, 'W', 13-50, 'T', 52-84, 'G', 86-120, 'L', 121-130, 'I', 132-282, 'S', 284-571, 'Q',
 A:Cross-references: EMBL:W92089; NID:G164613; PDB:AAA31102.1; PDB:G164614
 A:Experimental source: mammary gland
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-703/Product: lactoferrin #status predicted <MAT>
 F:20-350/Domain: transferrin repeat homology <TRH1>
 F:35-48/Region: antimicrobial
 F:354-691/Domain: transferrin repeat homology <TRH2>
 F:28-62,38-53,129-212,171-187,184-195,245-259,362-384,372-385,419-698,439-661,471-546,49
 F:77,107,206,267/Binding site: iron (Asp, Tyr, Tyr, His) #status predicted
 F:135/Binding site: carbonate (Arg) #status predicted
 F:409,447,540,609/Binding site: iron (Asp, Tyr, Tyr, His) #status predicted
 F:477/Binding site: carbonate (Arg) #status predicted
 F:490/Binding site: carbohydrate (Asn) (covalent) #status predicted

A/Cross-references: GB:AL51382; PIDN:CAD05733.1; PID:g16503226; GSPDB:GN00176
 A/Genetics: STY2193
 A/Genes: SLY2193
 C/Superfamily: Pyrococcus horikoshii hypothetical protein PH0926

Query Match 35.4%; Score 58; DB 2; Length 271;
 Best Local Similarity 35.7%; Pred. No. 0.71;
 Matches 10; Conservative 4; Mismatches 14; Indels 0; Gaps 0;
 QY 2 EWSKCYQWRMRKLGAPSTTCIRRTSA 29
 DB 25 EWQPAAPWLTRLHSGVPIILCSKXTAA 52

RESULT 12

TFHUM

N/Alternate names: melanoma-associated antigen gp95/p97
 C/Species: Homo sapiens (man)
 C/Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 28-Jan-2000
 C/Accession: A23814; A60925
 R/Rose, T.M.; Plowman, G.D.; Teplow, D.B.; Dreyer, W.J.; Hellstrom, K.E.; Brown, J.P.
 Proc. Natl. Acad. Sci. U.S.A. 83, 1261-1265, 1986
 A/Title: Primary structure of the human melanoma-associated antigen p97 (melanotransferin)
 A/Reference number: A23814; MUID:86149285; PMID:2419904
 A/Accession: A23814
 A/Molecule type: mRNA
 A/Residues: 1-738 <ROS>
 A/Cross-references: EMBL:M12154; NID:g189515; PIDN:AAA5992.1; PID:g189518
 A/Experimental source: melanoma
 R/Furukawa, K.S.; Furukawa, K.; Reai, F.X.; Old, L.J.; Lloyd, K.O.
 J. Exp. Med. 169, 585-590, 1989
 A/Title: A unique antigenic epitope of human melanoma is carried on the common melanoma
 A/Reference number: A60925; MUID:89094252; PMID:2463331
 A/Accession: A60925
 A/Molecule type: protein
 A/Residues: 20-25, 'X', 27-28, 'X', 30 <FUR>
 C/Comment: This protein is found predominantly in human melanomas and in certain fetal C/Genetics: Seven disulfide bonds are predicted in each domain.
 A/Genes: GDB:MFI2
 A/Cross-references: GDB:119387; OMIM:155750
 A/Map position: 3q28-3q29
 C/Superfamily: transferrin, transferrin repeat homology
 C/Keywords: blocked carboxyl end; duplication; glycoprotein; lipoprotein; membrane prote
 F/1-19/Domain: signal sequence #status predicted <SIG>
 F/18-360/Domain: transferrin repeat homology <TRH>
 F/20-710/Product: melanotransferrin #status predicted <MTF>
 F/361-709/Domain: transferrin repeat homology <TRH2>
 F/711-738/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F/38,135,515/Binding site: carbonyl (Asn) (covalent) #status predicted
 F/710/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature form

Query Match 35.1%; Score 57.5; DB 1; Length 738;
 Best Local Similarity 43.3%; Pred. No. 2.2;
 Matches 13; Conservative 3; Mismatches 13; Indels 1; Gaps 1;
 QY 1 PEWSKCYQWRMRKLGAPSTTCIRRTSA 29
 DB 31 PEQHKCQNSSEAFREAGIOPSLLCVGRGTA 60

RESULT 13

A47353

phosphatidylethanolamine N-methyltransferase (EC 2.1.1.17), 20K - rat
 C/Species: Rattus norvegicus (Norway rat)
 C/Date: 08-May-1995 #sequence_revision 12-May-1995 #text_change 05-May-2000
 C/Accession: A47353

R/Cui, Z.; Vance, J.E.; Chen, M.H.; Voelker, D.R.; Vance, D.E.
 J. Biol. Chem. 268, 16655-16663, 1993

A/Title: Cloning and expression of a novel phosphatidylethanolamine N-methyltransferase.
 A/Reference number: A47353; MUID:93346418; PMID:8344945
 A/Accession: A47353

A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-199 <CUI>
 A/Cross-references: GB:IL4441; NID:g310194; PIDN:AAA03154.1; PID:g310195
 C/Superfamily: phosphatidyl-N-methyltransferase, S-adenosylmethionine
 C/Keywords: membrane protein; methyltransferase; S-adenosylmethionine

Query Match 31.1%; Score 51; DB 2; Length 199;
 Best Local Similarity 36.0%; Pred. No. 5.6;
 Matches 9; Conservative 6; Mismatches 6; Indels 4; Gaps 1;
 QY 3 WSKCYQWRMRKLT---GAPSTTC 23
 DB 29 WNVVAREQRTKLSRAFGSPYLAC 53

RESULT 14

C70655

probable monooxygenase - Mycobacterium tuberculosis (strain H37Rv)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C/Accession: C70655

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 ; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: C70655

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA
 A/Residues: 1-489 <COL>
 A/Cross-references: GB:Z83864; GB:AL123456; NID:g3261687; PIDN:CA06212.1; PID:g301250;

A/Experimental source: strain H37Rv

C/Genetics: A/Genes: RV3854C

Query Match 31.1%; Score 51; DB 2; Length 489;
 Best Local Similarity 57.1%; Pred. No. 13;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 3 WSKCYQWRMRKLT 16
 DB 250 YSACQKMPRRMKM 263

RESULT 15

S33761

transferrin precursor - horse

N/Alternate names: growth-promoting factor

C/Species: Equus caballus (domestic horse)

C/Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999

C/Accession: S33761; S02145

R/Carpenter, M.A.; Broad, T.E.

Biochim. Biophys. Acta 1173, 230-232, 1993

A/Title: The cDNA sequence of horse transferrin.

A/Reference number: S33761; MUID:93277958; PMID:8504171

A/Accession: S33761

A/Molecule type: mRNA

A/Residues: 1-706 <CAR>

A/Cross-references: EMBL:M69020; NID:g164242; PIDN:AAA30958.1; PID:g164243

A/Experimental source: liver; developmental stage adult

R/Toshinari, K.; Yuasa, K.; Iga, F.; Yamura, A.

Biochim. Biophys. Acta 1010, 28-34, 1989

A/Title: A growth-promoting factor for human myeloid leukemia cells from horse serum idc

A/Reference number: S02145; MUID:89076897; PMID:2909248

A/Accession: S02145

A/Molecule type: protein

A/Residues: 20-35, 'X', 37, 'X', 39-40, 'X', 43-44 <YOS>

C/Complex: monomer

C/Superfamily: transferrin; transferrin repeat homology

C/Keywords: duplication; glycoprotein; iron transport; metal binding; plasma

F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-706/Product: transferrin #status experimental <MAT>
F;358-694/Domain: transferrin repeat homology <TRH2>
F;26-64,36-55,134-215,174-190,177-198,187-200,248-262,360-623,366-398,376-389,423-701,44

Query Match 30.5%; Score 50; DB 2; Length 706;
Best Local Similarity 34.5%; Pred. No. 26;
Matches 10; Conservative 7; Mismatches 10; Indels 2; Gaps 1;

QY 2 EMSKCYOWORRMRL--GAPSTICIRRTS 28
| | | | | : : : : :
Db 32 EVSKCASFRDMSKSIVPAPPLVACVKRTS 60

Search completed: July 30, 2003, 16:31:17
Job time : 18.4702 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 57.1369 Seconds

(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-6

Perfect score: 164

Sequence: 1 PEMSKCYQMGRMRRLGAPSTICRTSTA 29

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: A_Geneseq_13Jun03.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	164	100.0	29	21	AAV88341	Peptide derived fr
2	164	100.0	708	20	AA680774	Goat lactoferrin-as
3	163	99.4	708	20	AA680775	Goat lactoferrin-as
4	136	82.9	25	21	AAV83112	Synthetic lactofer
5	136	82.9	25	21	AAV85013	Caprine lactoferrin
6	122	74.4	689	21	AAV49270	Lactoferrin sequen
7	122	74.4	708	17	AAW09343	Bovine lactoferrin
8	122	74.4	708	19	AAW57318	Bovine lactoferrin
9	122	74.4	708	20	AAW6022	Bovine lactoferrin

10	122	74.4	708	21	AAW08183	Amino acid sequenc
11	122	74.4	708	22	AA664827	Chronic hepatitis
12	122	74.4	708	22	AAE02342	Bovine lactoferrin
13	120	73.2	32	14	AAV44845	Lactoferrin-relate
14	120	73.2	32	15	AAV48534	Lactoferrin-relate
15	120	73.2	32	15	AAV57466	Lactoferrin derive
16	120	73.2	32	16	AAV84703	Bovine lactoferrin
17	120	73.2	32	16	AAV80268	Anti-parasitic lac
18	120	73.2	32	17	AAV91856	Lactoferrin-derive
19	120	73.2	32	17	AAV90610	Lactoferrin-derive
20	120	73.2	32	17	AAV87625	Lactoferrin-derive
21	120	73.2	32	18	AAV26154	Lactoferrin deriva
22	120	73.2	32	18	AAV14040	Anti-parasitic pep
23	120	73.2	32	19	AAV70314	Thrombus formation
24	120	73.2	32	19	AAV53229	Lactoferrin hydrol
25	120	73.2	32	19	AAV41294	Apoptosis inducer
26	120	73.2	32	21	AAV55543	Peptide SEQ ID NO:
27	120	73.2	38	14	AAV48444	Lactoferrin-relate
28	120	73.2	38	15	AAV48533	Lactoferrin derive
29	120	73.2	38	15	AAV57465	Lactoferrin derive
30	120	73.2	38	16	AAV84702	Bovine lactoferrin
31	120	73.2	38	16	AAV80267	Anti-parasitic lac
32	120	73.2	38	17	AAV91855	Lactoferrin-derive
33	120	73.2	38	17	AAV90609	Lactoferrin derive
34	120	73.2	38	17	AAV87624	Lactoferrin deriva
35	120	73.2	38	18	AAV26153	Lactoferrin decomp
36	120	73.2	38	18	AAV14039	Anti-parasitic pep
37	120	73.2	38	19	AAV70313	Thrombus formation
38	120	73.2	38	19	AAV53228	Lactoferrin hydrol
39	120	73.2	38	19	AAV41293	Apoptosis inducer
40	120	73.2	38	21	AAV55542	Peptide SEQ ID NO:
41	115	70.1	704	22	AAV51535	Recombinant lactof
42	105	64.0	51	17	AAV91194	Lactoferrin decomp
43	105	64.0	52	16	AAV87909	Bovine lactoferrin
44	100	61.0	36	14	AAV44846	Lactoferrin-relate
45	100	61.0	36	17	AAV10517	Lactoferrin derive

ALIGNMENTS

RESULT 1	
ID	AAV88341 strand; peptide; 29 AA.
AC	AAV88341;
XX	
DT	14-JUL-2000 (first entry)
DE	Peptide derived from an internal fragment of goat whey.
DE	
KW	Peptide production; biological fluid; milk; whey; blood;
KW	antibacterial peptide; lactoferrin; antiviral; antitumour activity.
XX	
OS	Capra sp.
XX	
FH	Key
FT	Disulfide-bond 6.23 Location/Qualifiers
FT	/note="Optionally not present"
XX	
XX	WO200015655-A1.
XX	
PD	.23-MAR-2000.
XX	
PF	15-SEP-1999; 99WO-BP07002.
XX	
PR	15-SEP-1998; 98EP-0203107.
PR	08-JUN-1999; 99EP-0201815.
XX	
PA	(NIZO-) NIZO FOOD RES.
XX	
PI	Visser S, Reclio I;
XX	

DR WPI; 2000-271377/23.

XX Novel process for producing peptides with e.g. antimicrobial activity
PT from biological fluids such as milk, whey or blood comprises contacting
PT fluid with chromatographic medium to adsorb peptide domain of interest

PS Claim 14; Page 22; 41pp; English.

CC This sequence represents a peptide derived from an internal fragment of
CC goat whey. The peptide is an example of a peptide with antibacterial
CC activity that can be produced by the process of the invention. The
CC invention relates to a process for producing peptides from biological
CC fluids. The process comprises chromatography of the biological fluid, in
CC situ hydrolysis of selectively bound peptides, washing to remove unbound
CC peptide, and elution of the peptides of interest. The process is used for
CC producing peptides from biological fluids, such as milk, whey or blood.
CC For example, the process can be used to produce antibacterial peptides
CC derived from lactoferrin, using cheese whey as a starting material. The
CC peptides obtained have preferably antimicrobial and/or antiviral and/or
CC antitumour activity. The process of the invention is relatively simple
CC and generally economically and technically more attractive than those
CC methods previously used. The method provides high yield peptides with a
CC purification of the precursor protein.

XX Sequence 29 AA;

Query Match 100.0%; Score 164; DB 21; Length 29;
Best Local Similarity 100.0%; Pred. No. 8.4e-17;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMSKCYQWQRRMRKLGAPSTICIRRTSA 29
DB 1 PMSKCYQWQRRMRKLGAPSTICIRRTSA 29

RESULT 2

AAG80774 AAG80774 standard; Protein; 708 AA.

AC AAG80774;

DT 19-APR-2002 (first entry)

DE Goat lactoferrin-associated protein #1.

XX Primer; lactoferrin; antibiotic; goat.

XX Capra sp.

OS Location/Qualifiers

PH Key Misc-difference 360 /note= "Encoded by CGG"

FT KR98043944-A.

XX 05-SEP-1998.

XX 05-DEC-1996; 96KR-0061925.

XX 05-DEC-1996; 96KR-0061925.

XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.

XX Yoo DY, Nam MS, Lee TH, Yoo SR, Kim SJ, Lee GG;

XX WPI; 1999-465631/39.

XX N-PSDB; ABA97187.

PT Antibiotic lactoferrin derived from Korean black goat and gene thereof -
PT NoAbstract

PS Disclosure; Page 6-9; 11pp; Korean.

XX This sequence represents the goat antibiotic lactoferrin described in the
CC disclosure of the invention.

XX Sequence 708 AA;

Query Match 100.0%; Score 164; DB 20; Length 708;
Best Local Similarity 100.0%; Pred. No. 2.6e-15;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMSKCYQWQRRMRKLGAPSTICIRRTSA 29
DB 33 PMSKCYQWQRRMRKLGAPSTICIRRTSA 61

RESULT 3

AAG80775 AAG80775 standard; Protein; 708 AA.

AC AAG80775;

DT 19-APR-2002 (first entry)

DE Goat lactoferrin-associated protein #2.

XX Primer; lactoferrin; antibiotic; goat.

XX Capra sp.

OS Location/Qualifiers

PH Key Misc-difference 360 /note= "Encoded by CGG"

FT KR98043944-A.

XX 05-SEP-1998.

XX 05-DEC-1996; 96KR-0061925.

XX 05-DEC-1996; 96KR-0061925.

XX (KOAD) KOREA ADV INST SCI & TECHNOLOGY.

XX Yoo DY, Nam MS, Lee TH, Yoo SR, Kim SJ, Lee GG;

XX WPI; 1999-465631/39.

XX Antibiotic lactoferrin derived from Korean black goat and gene thereof -
PT NoAbstract

XX Disclosure; Page 6-9; 11pp; Korean.

CC This sequence represents the goat antibiotic lactoferrin described in the
CC disclosure of the invention.

XX Sequence 708 AA;

Query Match 99.4%; Score 163; DB 20; Length 708;
Best Local Similarity 96.6%; Pred. No. 3.6e-15;
Matches 28; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PMSKCYQWQRRMRKLGAPSTICIRRTSA 29
DB 33 PMSKCYQWQRRMRKLGAPSTICIRRTSA 61

RESULT 4

AAV83112 AAV83112 standard; peptide; 25 AA.

XX AAV83112;

DT 24-JUL-2000 (first entry)
 XX Synthetic lactoferrin fragment.
 DE Lactoferrin; antimicrobial; cytostatic; infection; tumour;
 XX sterilisation; wound healing; spermicide; goat.
 KM Synthetic.
 OS Capra hircus.
 XX MO200012541-A2.
 PN 09-MAR-2000.
 PD 31-AUG-1999; 99WO-GB02850.
 PF 28-AUG-1998; 98GB-0018938.
 PR (ALPH-) ALPHARMA AS.
 PA (GARD/) GARDNER R.
 XX Svendsen JS, Rekdal O, Sveinbjornsson B, Vorland L;
 PI MPI, 2000-270793/23.
 DR Cytotoxic peptides useful as medicament for treating tumours and
 XX bacterial infections, comprises one or more non-genetic bulky and
 XX lipophilic amino acids
 XX Example 1; Figure 1; 114pp; English.
 PS Cytotoxic 7-25 mer lactoferrin peptides which comprise three or more
 CC cationic residues and have one or more non-genetic bulky and
 CC lipophilic amino acids have cytostatic and antimicrobial activity.
 CC The peptides are useful as medicament for treating bacterial
 CC infections and tumours. They are also useful for sterilising
 CC agents for materials susceptible to microbial contamination. They may
 CC also be used as promoters of wound healing and spermicides. The small
 CC size of the peptides gives them an increased half life and allows
 CC efficient biodelivery. The peptides can be administered without need
 CC for an injection, such as by inhalation or by absorption across the
 CC blood capillaries of the nasal passages. This synthetic lactoferrin
 CC peptide corresponds to amino acids 17-41 of caprine lactoferrin.
 CC
 XX SQ Sequence 25 AA;
 Query Match 82.9%; Score 136; DB 21; Length 25;
 Best Local Similarity 96.0%; Pred. No. 8.8e-13;
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 4 SKCYQWQRRMRKLGAPSTICIRRTS 28
 |||||
 DB 1 SKCYQWQRRMRKLGAPSTICIRRTS 25
 |||||
 RESULT 5
 ID AAY85013 standard; peptide; 25 AA.
 XX AAY85013;
 AC 20-JUN-2000 (first entry)
 DT Caprine lactoferrin peptide LfC(17-41) amino acid sequence.
 XX
 DB Lactoferrin; lactoferrin; antibacterial; bacterial infection;
 KM bacterial growth; solid tumour; treatment; goat.
 XX
 OS Capra sp.
 XX MO200012542-A2.
 PN 09-MAR-2000.
 PD

XX 31-AUG-1999; 99WO-GB02851.
 XX 28-AUG-1998; 98GB-0018938.
 XX (ALPH-) ALPHARMA AS.
 PA (GARD/) GARDNER R.
 XX Svendsen JS, Rekdal O, Sveinbjornsson B, Vorland L;
 PI MPI, 2000-256582/22.
 DR Modified cytotoxic lactoferrin peptide useful for preparing
 PT pharmaceutical composition against bacterial infections and tumours and
 PT for inhibiting bacterial growth
 XX Example 1; Fig 1; 112pp; English.
 PS This sequence represents a caprine lactoferrin peptide. Lactoferrin is
 CC a weak iron transporter which also shows weak antibacterial effect.
 CC Lactoferrin B (LFB) is a peptide produced when bovine lactoferrin is
 CC digested with pepsin. LFB is more active than lactoferrin. The two
 CC tryptophan residues at positions 6 and 8 in LFB 17-31 cannot be
 CC substituted with alanine without loss of antibacterial activity. The
 CC invention relates to a modified cytotoxic lactoferrin peptide of 7-25
 CC amino acids in length with three cationic residues and one or more extra
 CC bulky and lipophilic amino acids, than its native peptide. The invention
 CC also relates to the esters, amides, salts and cyclic derivatives of the
 CC modified peptide. The modified lactoferrin peptides of the invention are
 CC useful for treating bacterial infections and tumours, for preparing
 CC medicaments and for inhibiting bacterial growth. LFB and its fragments
 CC are useful for manufacturing medicaments for treating solid tumours. The
 CC modified peptides have a very small size, which is suitable for
 CC biodelivery. Due to the smaller size, the circulating half-life of the
 CC peptide is increased as they are less vulnerable to endopeptidases. The
 CC observed cytolytic effect of the modified lactoferrin peptides in tumours
 CC is not species specific and they have greater utility in treating human
 CC tumours.
 XX SQ Sequence 25 AA;
 Query Match 82.9%; Score 136; DB 21; Length 25;
 Best Local Similarity 96.0%; Pred. No. 8.8e-13;
 Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 4 SKCYQWQRRMRKLGAPSTICIRRTS 28
 |||||
 DB 1 SKCYQWQRRMRKLGAPSTICIRRTS 25
 |||||
 RESULT 6
 ID AAY49270 standard; protein; 689 AA.
 XX AAY49270;
 AC 07-FEB-2000 (first entry)
 DT Lactoferrin sequence.
 XX
 DE Adhesion inhibition; enteropathogenic; Escherichia coli; lactoferrin;
 KM medicament; drink; feed; food poisoning.
 XX
 OS Unidentified.
 XX
 PN JP11292789-A.
 PD 26-OCT-1999.
 PF 03-APR-1998; 98JP-0107167.
 PR 03-APR-1998; 98JP-0107167.
 XX

PA (MEIP) MEIJI MILK PROD CO LTD.
 XX
 DR WPI; 2000-018674/02.
 XX
 PT Adhesion inhibitory compositions of enteropathogenic E. coli to cells -
 PS used in medicines and foodstuffs
 XX
 PS Disclosure; Fig 4; 7pp; Japanese.
 CC The invention provides an adhesion inhibitory composition of
 CC enteropathogenic Escherichia coli to cells. The composition contains a
 CC lactoferrin degraded matter. The composition is useful as a medicament,
 CC a drink and food or a feed. The composition is effective for prevention
 CC and/or treatment of infectious food poisoning caused by enteropathogenic
 CC E. coli.
 XX
 SO Sequence 689 AA;
 Query Match 74.4%; Score 122; DB 21; Length 689;
 Best Local Similarity 72.4%; Pred. No. 3.4e-09;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 1 PEMSKCYOMORRMKLGAPSTICIRRTSA 29
 DB 14 PEMFKCRMQMRMKLGAPSTICVRRAPA 42
 RESULT 7
 ID AAM09343 standard; Protein; 708 AA.
 XX
 AC AAM09343;
 XX
 DT 25-MAR-2003 (updated)
 DT 18-MAR-1997 (first entry)
 XX
 DE Bovine lactoferrin.
 XX
 KM Human; lactoferrin; iron-binding glycoprotein; milk; secretion; fungus;
 KM transferrin; bactericidal activity; prostate; expression system; primer;
 KM PR; polymerase chain reaction; amplification; signal peptide; antiviral;
 KM alpha-amylase; Aspergillus oryzae; nutrition; bovine.
 OS
 OS Bos taurus.
 XX
 PN USS571691-A.
 XX
 PD 05-NOV-1996.
 XX
 PF 28-OCT-1993; 93US-0145681.
 XX
 PR 28-OCT-1993; 93US-0145681.
 PR 05-MAY-1989; 89US-0348270.
 PR 28-SEP-1989; 89US-0413880.
 PR 24-APR-1992; 92US-0873304.
 PR 27-OCT-1992; 92US-0967947.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX
 PI Conneely OM, Headon DR, May GS, O'Malley BW;
 XX
 DR WPI; 1996-505400/50.
 DR N-PSDB; AAT48031.
 XX
 PT New DNA encoding human lactoferrin, its natural alleles and
 PT substitution analogues - useful e.g. for preventing iron deficiency
 PT and as antiviral/antimicrobial agent
 XX
 PS Disclosure; Column 31-34; 92pp; English.
 CC This is the amino acid sequence of bovine lactoferrin, an iron-binding
 CC glycoprotein found in milk and other secretions and body fluids.
 CC Fragments of the protein are also known to have biological activity e.g.

CC the N-terminal portion of the protein has a bactericidal activity. The
 CC gene was used to construct a fusion protein in which the native
 CC lactoferrin signal peptide was replaced by the alpha-amylase II signal
 CC peptide. The novel construct was then expressed in Aspergillus oryzae.
 CC The protein can be used for antibacterial and antiviral activities as
 CC well as an iron-carrying protein for nutritional or therapeutic
 CC applications.
 CC (updated on 25-MAR-2003 to correct PF field.)
 XX
 SO Sequence 708 AA;
 Query Match 74.4%; Score 122; DB 17; Length 708;
 Best Local Similarity 72.4%; Pred. No. 3.5e-09;
 Matches 21; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 1 PEMSKCYOMORRMKLGAPSTICIRRTSA 29
 DB 33 PEMFKCRMQMRMKLGAPSTICVRRAPA 61
 RESULT 8
 ID AAM57318 standard; Protein; 708 AA.
 XX
 AC AAM57318;
 XX
 DT 10-AUG-1998 (first entry)
 DT
 XX
 DE Bovine lactoferrin.
 XX
 KM Bovine; lactoferrin; recombinant; therapeutic; nutritional; iron;
 KM Fe binding site; bacteria; bactericidal; milk.
 XX
 OS
 OS Bos taurus.
 XX
 PN USS766939-A.
 XX
 PD 16-JUN-1998.
 XX
 PF 30-MAY-1995; 95US-0453703.
 XX
 PR 28-OCT-1993; 93US-0145681.
 PR 05-MAY-1989; 89US-0348270.
 PR 24-APR-1992; 92US-0873304.
 PR 27-OCT-1992; 92US-0967947.
 PR 30-MAY-1995; 95US-0453703.
 XX
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 XX
 PI Conneely OM, Headon DR, May GS, O'Malley BW;
 XX
 DR WPI; 1998-361705/31.
 DR N-PSDB; AAV30771.
 XX
 PT Plasmids containing human lactoferrin DNA - for recombinant
 PT production of the enzyme, especially fragments having bactericidal
 PT activity
 XX
 PS Example 13; Fig 14; 92pp; English.
 CC The present sequence represents bovine lactoferrin. A plasmid has been
 CC developed which is suitable for the expression of a human lactoferrin,
 CC or an iron-binding lobe of lactoferrin, in a transformed prokaryotic
 CC host cell. The plasmid comprises a DNA sequence encoding a naturally
 CC occurring human lactoferrin protein or an iron-binding lobe of
 CC lactoferrin and further comprises transcriptional and translational
 CC regulatory elements capable of regulating the expression of the
 CC lactoferrin-encoding DNA sequence in the transformed host cell. The
 CC plasmid is useful for producing recombinant human lactoferrin proteins
 CC in bacteria. Lactoferrin is a 78 kDa iron-binding glycoprotein found
 CC in milk and other secretory fluids. It is involved in iron transfer
 CC and delivery in mammals. It has been implicated as a resistance
 CC factor in suckled new born infants against enteritis infections; the

XX Lactoferrin; antibiotic; chelate; mastitis; bowel disorder;
 KW disease; bacteria; yeast; fungi; disinfection; drug; foodstuff;
 XX cosmetic; toiletries.
 XX Unidentified.
 OS
 XX
 FH Key Location/Qualifiers
 FT Disulfide bond 10..27
 XX
 XX MO9314640-A1.
 PN
 XX
 PD 05-AUG-1993.
 XX
 XX 30-NOV-1992; 92WO-JP01563.
 PF
 XX 23-JUN-1992; 92JP-0032660.
 PR 11-MAR-1992; 92JP-0052943.
 PR 30-SEP-1992; 92JP-0262143.
 PR 30-SEP-1992; 92JP-0262559.
 XX
 PA (MORG) MORINAGA MILK IND CO LTD.
 XX
 XX Bellamy W, Fukuwatari Y, Kawase K, Shimamura S;
 PI Takase M, Tokitay, Tomita M, Wakabayashi H, Yamauchi K;
 DR WPI; 1993-258265/32.
 XX
 XX Antibacterial agent comprising decomposition products of
 PT lactoferrin - with chelate e.g. EDTA alcohol and/or antibiotic
 PT e.g. penicillin, also useful against yeast and fungi
 XX
 PS Disclosure; Page 87; 100pp; Japanese.
 XX
 CC Lactoferrin-related peptides are used in new antibacterial compsn.
 CC The compsn. is highly effective against a broad range of bacteria,
 CC yeasts and fungi. It can be used therapeutically (internal and
 CC external application), e.g. for mastitis, bowel disorders, urinary
 CC infections, etc. It can also be used for the disinfection and
 CC protection of drugs, foodstuffs, cosmetics and toiletries and
 CC household items (such as kitchen towels and toilet paper).
 CC (Updated on 09-JAN-2003 to add missing OS field.)
 CC
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 XX
 SQ Sequence 32 AA;
 XX
 Query Match 73.2%; Score 120; DB 14; Length 32;
 Best Local Similarity 76.9%; Pred. No. 2.5e-10;
 Matches 20; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 PEMSKCYQWORMRKIGAPSIICRR 26
 DB 5 PEMFKCRWQMVKKIGAPSIICVR 30

RESULT 14
 AAR48534
 ID AAR48534 standard; peptide; 32 AA.
 XX
 AC AAR48534;
 XX
 DT 25-MAR-2003 (updated)
 DT 10-AUG-1994 (first entry)
 XX
 DE Lactoferrin derived peptide #28.
 XX
 KW Decomposition; lactoferrin; digestion; enzyme; pepsin; trypsin;
 KW antioxidant; oxidation; inhibitor; vitamin B; ascorbic acid;
 KW vitamin A; beta-carotene; superoxide dismutase; coenzyme Q;
 KW lipid oxidation; foodstuff; drugs; health food; toiletries; cosmetics.
 XX
 OS Bos taurus.
 XX

PN MO9403555-A1.
 XX
 PD 17-FEB-1994.
 XX
 PF 04-AUG-1993; 93WO-JP01090.
 XX
 PR 07-AUG-1992; 92JP-0211335.
 XX
 PA (MORG) MORINAGA MILK IND CO LTD.
 XX
 PI Bellamy WR, Fukuwatari Y, Kawase K, Shimamura S;
 PI Takase M, Tokitay, Tomita M, Wakabayashi H, Yamauchi K;
 DR WPI; 1994-065650/08.
 XX
 XX Antioxidant peptide lactoferrin decomposition product - prevents
 PT oxidation of lipid(s) in foodstuffs and drugs without affecting
 PT their taste
 XX
 PS Claim 3; Page 39; 47pp; Japanese.
 XX
 CC The sequences given in AAR48507-37 are peptides derived by the
 CC decomposition of lactoferrin, pref. by digestion with an enzyme, eg.
 CC pepsin or trypsin. These peptides may be used in an antioxidant
 CC composition which may also contain an oxidation inhibitor such as
 CC vitamin B, ascorbic acid, vitamin A, beta-carotene, superoxide
 CC dismutase or coenzyme Q. The antioxidant prevents lipid oxidation
 CC in foodstuffs, drugs, health foods, toiletries and cosmetics.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 XX
 SQ Sequence 32 AA;
 XX
 Query Match 73.2%; Score 120; DB 15; Length 32;
 Best Local Similarity 76.9%; Pred. No. 2.5e-10;
 Matches 20; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 PEMSKCYQWORMRKIGAPSIICRR 26
 DB 5 PEMFKCRWQMVKKIGAPSIICVR 30

RESULT 15
 AAR57466
 ID AAR57466 standard; protein; 32 AA.
 XX
 AC AAR57466;
 XX
 DT 28-FEB-1995 (first entry)
 DT XX
 DE Lactoferrin derived peptide #29.
 XX
 KW Lactoferrin; chemical; enzymatic; hydrolysis; antimicrobial;
 KW antiseptic; ischaemic disease.
 KW
 XX Mus musculus.
 OS
 XX
 PN JP06172200-A.
 XX
 PD 21-JUN-1994.
 PD
 PF 08-DEC-1992; 92JP-0327738.
 PF
 PR 08-DEC-1992; 92JP-0327738.
 PR
 XX (MORG) MORINAGA MILK IND CO LTD.
 PA
 PA WPI; 1994-238662/29.
 XX
 DR Brain protectant for preventing ischaemic diseases without side
 PT effects - comprising 31 specified peptide(s), prepd. by
 PT lactoferrin hydrolysis
 XX
 PS Disclosure; Page 10; 11pp; Japanese.
 XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 27.1429 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-7

Perfect score: 82
Sequence: 1 TORKTRNGFRVPLARE 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvrius:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	712	6 Q9MZY2	Q9MZY2 ovis aries
2	55	67.1	711	6 Q9GJW6	Q9GJW6 Camelus dromedarius
3	45	54.9	710	11 Q8F481	Q8F481 mesocricetus
4	44	53.7	359	16 Q9KXV1	Q9KXV1 vibrio cholerae
5	44	53.7	667	3 Q8XOK1	Q8XOK1 neurospora crassa
6	43	52.4	725	13 Q8OEX3	Q8OEX3 xenopus laevis
7	43	52.4	1134	16 Q9RTU0	Q9RTU0 streptomyces
8	42	51.2	710	11 Q9JWAO	Q9JWAO mus musculus
9	42	51.2	1394	16 Q8G3M2	Q8G3M2 bifidobacterium
10	41	50.0	138	16 Q9PF48	Q9PF48 xylophaga
11	41	50.0	414	5 Q21764	Q21764 caenorhabditis elegans
12	41	50.0	417	16 Q8FLU8	Q8FLU8 corynebacterium
13	41	50.0	631	16 Q912G3	Q912G3 pseudomonas
14	41	50.0	876	10 Q64483	Q64483 arabidopsis
15	40	48.8	1333	16 Q8NT25	Q8NT25 corynebacterium
16	40	48.8	1333	16 Q8FS96	Q8FS96 corynebacterium

17	39.5	48.2	1171	16 Q92PS9	Q92PS9 rhizobium m
18	39	47.6	76	4 Q9YJL1	Q9YJL1 homo sapien
19	39	47.6	95	15 Q9ONT7	Q9ONT7 human immun
20	39	47.6	126	2 Q8GJK8	Q8GJK8 synecococc
21	39	47.6	128	17 Q59561	Q59561 pyrococcus
22	39	47.6	182	5 Q9N439	Q9N439 caenorhabd
23	39	47.6	226	16 Q8KF28	Q8KF28 chlorobium
24	39	47.6	235	10 Q9C6S3	Q9C6S3 arabidopsis
25	39	47.6	314	10 Q8VZT8	Q8VZT8 arabidopsis
26	39	47.6	320	5 Q9SU84	Q9SU84 drosophila
27	39	47.6	352	17 Q97917	Q97917 thermoplasma
28	39	47.6	376	16 Q8XKR9	Q8XKR9 raietonia b
29	39	47.6	418	16 Q8EL08	Q8EL08 oceanobacill
30	39	47.6	625	4 Q8N4A6	Q8N4A6 homo sapien
31	39	47.6	629	16 Q8Z050	Q8Z050 anabaena sp
32	39	47.6	656	4 Q8WV21	Q8WV21 homo sapien
33	39	47.6	684	10 Q9ZV55	Q9ZV55 arabidopsis
34	39	47.6	1029	4 Q8TBX2	Q8TBX2 homo sapien
35	39	47.6	1156	4 Q8IUL8	Q8IUL8 homo sapien
36	39	47.6	1281	12 Q85425	Q85425 rat cytochrome
37	39	47.6	127	4 Q9UP09	Q9UP09 homo sapien
38	38.5	47.0	310	17 Q8UIF1	Q8UIF1 pyrococcus
39	38	46.3	123	16 Q8Y118	Q8Y118 anabaena sp
40	38	46.3	172	16 P96883	P96883 mycobacteri
41	38	46.3	185	16 Q9X8S3	Q9X8S3 streptomyce
42	38	46.3	204	4 Q9NTP6	Q9NTP6 homo sapien
43	38	46.3	251	2 Q9AMH3	Q9AMH3 streptomyce
44	38	46.3	251	2 Q9PSJ0	Q9PSJ0 streptomyce
45	38	46.3	268	17 Q8TTV3	Q8TTV3 methanosaarc

ALIGNMENTS

RESULT 1

Q9MZY2 PRELIMINARY; PRT; 712 AA.
AC Q9MZY2; 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Airway lactoperoxidase.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Tracheal mucosa;
RX MEDLINE=20298390; PubMed=10837362;
RA Gerson C., Sabater J., Scuti M., Torbatti A., Coffey R., Abraham J.W.,
RA Lauredo I., Forteza R., Wanner A., Salathe M., Abraham W., Conner G.,
RT "The Lactoperoxidase System Functions in Bacterial Clearance of Airways.";
RL Am. J. Respir. Cell Mol. Biol. 22:665-671 (2000).
DR EMBL; AF027970; AAF74979.1; -.
DR HSSP; P05164; ICXP.
DR InterPro; IPR002007; Anim_peroxidase.
DR Pfam; PF03098; An_peroxidase; 1.
DR PRINTS; PR00457; AMPEROXIDASE.
KW Peroxidase.
SQ SEQUENCE 712 AA; 80486 MW; 5D3465943D6426F2 CRC64;

Query Match 100.0%; Score 82; DB 6; Length 712;
Best Local Similarity 100.0%; Pred. No. 4.5e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TORKTRNGFRVPLARE 16
DB 179 TORKTRNGFRVPLARE 194

RESULT 2

Q9GJW6 PRELIMINARY; PRT; 711 AA.
 AC Q9GJW6, 16, Created
 DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Peroxidase precursor (EC 1.11.1.7).
 GN LACTOPEROXIDASE.
 OS Camelus dromedarius (Dromedary) (Arabian camel).
 OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 NCBI_TaxID=9838;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lactating mammary gland;
 RA Kappeler S.R., Farah Z., Fuhon Z.;
 RT "Sequence Analysis of Camel (Camelus dromedarius) Lactoperoxidase."
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ131675; CAC19552.1; -
 DR HSSP; P05164; IXP.
 DR InterPro; IPR002007; Anim_peroxidase.
 DR Pfam; PF03098; An_peroxidase; 1.
 DR PRINTS; PR00457; ANPEROXIDASE.
 KW Oxidoreductase; Peroxidase; Signal.
 FT SIGNAL 1 99
 FT CHAIN 100 711
 FT CHAIN PEROXIDASE.
 SQ SEQUENCE 711 AA; 80674 MW; 4E3328379BD750C5 CRC64;

Query Match

Best Local Similarity 67.1%; Score 55; DB 6; Length 711;
 Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TORKTRNGFRVPLARE 16
 DB 178 TRGKXKNGFPLPLARE 193

RESULT 3

Q8R481 PRELIMINARY; PRT; 710 AA.
 AC Q8R481, 21, Created
 DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Lactoperoxidase.
 GN LPO.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lactating mammary gland;
 RA Paliwal A., Srikanan S., De P.K.;
 RT "DNA Cloning and Regulation of a Lactoperoxidase (LPO) from Hamster Lactimal Gland."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF498045; AAM15535.1; -
 DR InterPro; IPR002007; Anim_peroxidase.
 DR Pfam; PF03098; An_peroxidase; 1.
 KW Peroxidase.
 SQ SEQUENCE 710 AA; 80218 MW; AD8C9359DB1C270F CRC64;

Query Match

Best Local Similarity 54.9%; Score 45; DB 11; Length 710;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 TORKTRNGFRVPLARE 16
 DB 177 TPGRTRNGFPLPLPOPRD 192

RESULT 4

Q9KXV1 PRELIMINARY; PRT; 359 AA.
 AC Q9KXV1, 15, Created
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE GDEF family protein.
 GN VCA0217.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; OC Vibrionaceae; Vibrio.
 NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=El Tor N16961 / Serotype O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bacs S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Uitterback T., Fleischmann R.D., Nieman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Frazer C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae."
 RL Nature 406:477-483 (2000).
 DR EMBL; AE004361; AAF96129.1; -
 DR TIGR; VCA0217; -
 DR InterPro; IPR000160; GDEF.
 DR Pfam; PF00990; GDEF; 1.
 DR SMART; SM00267; DUF1; 1.
 DR TIGRFAIR; TIGR00254; GDEF; 1.
 KW Complete proteome.
 SQ SEQUENCE 359 AA; 40583 MW; 6595577060520259 CRC64;

Query Match

Best Local Similarity 53.7%; Score 44; DB 16; Length 359;
 Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 OKRTNGFRVPLA 14
 DB 74 OKRTNGFRVPLA 86

RESULT 5

Q8XOK1 PRELIMINARY; PRT; 667 AA.
 AC Q8XOK1, 20, Created
 DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)
 DE Related to POL12 (DNA-directed DNA polymerase alpha).
 GN B1H24.020.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospora.
 NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Allyn V., Honeisiel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mennhant G.;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL670005; CAD21272.1; -
 DR Pfam; PF04058; DNA_pol_alpha_B; 1.
 KW DNA-directed DNA polymerase.
 SQ SEQUENCE 667 AA; 73322 MW; 4A23A5DE72D645C CRC64;

```

Query Match          53.7%; Score 44; DB 3; Length 667;
Best Local Similarity 69.2%; Pred. No. 28;
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 TOKTRNGRPVPL 13
Db 284 TSKKUGGGRVPL 296

RESULT 6
O8GFX3 PRELIMINARY; PRT; 725 AA.
AC O8GFX3;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Peroxidase 2'.
GN FOX2'.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith S.J., Koticha S., Towers N., Mohun T.J.;
RT "XLRP-1 and FOX2 peroxidase expression reveal the site of embryonic
RT myeloperoxidase in Xenopus.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; A1069942; AAL55400.1; -.
DR InterPro; IPR002007; Anim_peroxidase.
DR Pfam; PF03098; An_peroxidase; 1.
SQ SEQUENCE 725 AA; 82417 MW; 8546878368C4947F CRC64;

Query Match          52.4%; Score 43; DB 13; Length 725;
Best Local Similarity 53.3%; Pred. No. 46;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 TOKTRNGRPVPLAR 15
Db 187 TENRRNGPMLPLAR 201

RESULT 7
O9RJU0 PRELIMINARY; PRT; 1134 AA.
AC O9RJU0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Possible DNA-binding protein.
GN SC00370 OR SCF41.29C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleeser H.M., Denapate D., Richner A., Callum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for

```

```

RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleeser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleeser T., Larke J., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)";
RL Nature 417:141-147(2002).
DR EMBL; AL939105; CAB55732.1; -.
DR InterPro; IPR006311; Tat.
DR TIGRFAMs; TIGR01409; Tat_signal_seq; 1.
KW DNA-binding; Complete proteome.
SQ SEQUENCE 1134 AA; 121689 MW; 89B689CD455C4520 CRC64;

Query Match          52.4%; Score 43; DB 16; Length 1134;
Best Local Similarity 64.3%; Pred. No. 72;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 RTRNGRPVPLARE 16
Db 913 RATKNGRVPLDLE 926

RESULT 8
O9IWA0 PRELIMINARY; PRT; 710 AA.
AC O9IWA0;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical 79.9 kDa protein.
GN LPO.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Salivary gland;
RA Strausberg R.;
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016212; AAH16212.1; -.
DR WGD; WGI:1923363; LPO.
DR InterPro; IPR002007; Anim_peroxidase.
DR Pfam; PF03098; An_peroxidase; 1.
DR PRINTS; PR00457; ANPEROXIDASE.
KW Hypothetical protein.
SQ SEQUENCE 710 AA; 79922 MW; EB9BCF12472EC5DA CRC64;

Query Match          51.2%; Score 42; DB 11; Length 710;
Best Local Similarity 56.2%; Pred. No. 67;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 TOKTRNGRPVPLARE 16
Db 177 TPGKNGRPVPLQPRE 192

RESULT 9
O8G3M2 PRELIMINARY; PRT; 1394 AA.
AC O8G3M2;
DT 01-MAR-2003 (TREMBlrel. 23, Created)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)
DB Possible helicasae.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RA MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karimiantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwaan M.-C., Desiere F., Bork P., Delley M.,
RA Pridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AB014807; AAN25520.1; -.
KW Helicase; Complete proteome.
SQ SEQUENCE 1394 AA; 151986 MW; 7393F1750312D9F3 CRC64;

Query Match
Best Local Similarity 51.2%; Score 42; DB 16; Length 1394;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 4 KTRNGFRVPLARE 16
Db 694 QTVNGFRTPARK 706

RESULT 10
Q9PF48 PRELIMINARY; PRT; 138 AA.
AC Q9PF48;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE Hypothetical protein Xf0830.
GN Xf0830.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C; PubMed=10910347;
RA MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.B., Baia G.S., Baptista C.S.,
RA Barrios M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carver H.,
RA Colauto N.B., Colombo C., Costa F.P., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.V., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinelli A.P., Ferreira A.J.S., Ferreira V.C.A., Farto J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Gardier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.B., Kuramae E.R., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.B.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,

```

```

RA Zago M.A., Zatz M., Melandis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE003922; AAF83640.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 138 AA; 15117 MW; 1A9397C765730FC1 CRC64;

Query Match
Best Local Similarity 50.0%; Score 41; DB 16; Length 138;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 QKTRNGFRVPLAR 15
Db 123 QKTRDGRAPGAR 136

RESULT 11
Q21764 PRELIMINARY; PRT; 414 AA.
AC Q21764;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)
DE R05H5.4 protein.
GN R05H5.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurtry A.A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 287:2012-2018(1998).
DR EMBL; Z48795; CAA88727.1; -.
DR HSSP; P12931; IPMK.
DR WormPep; R05H5.4; CE02291.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR000980; SH2.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF00017; SH2; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Prot_kinase; 1.
DR SMART; SM00252; SH2; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 414 AA; 47188 MW; E51F5CD772E9A908 CRC64;

Query Match
Best Local Similarity 50.0%; Score 41; DB 5; Length 414;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 2 QKTRNGFRVPL 12
Db 327 KKKTRAGTRLP 337

RESULT 12
Q8FLU8 PRELIMINARY; PRT; 417 AA.
AC Q8FLU8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)

```


DT 01-MAR-2003 (TRENBLrel. 23, last annotation update)
 DE Putative L-lactate dehydrogenase.
 GN CE2762.
 OS Corynebacterium efficiens.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 OK NCBI_TaxID=152794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
 RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
 RA Ikeo K., Suzuki M., Mashima T., Itoh T., Yamagishi A., Nishio Y.,
 RA Usuda Y., Sugimoto S.;
 RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF005223; BAC19572.1; -.
 KM Complete proteome.
 SQ SEQUENCE 417 AA; 45519 MW; 127DD710E3C7A500 CRC64;

Query Match 50.0%; Score 41; DB 16; Length 417;
 Best Local Similarity 70.0%; Pred. No. 59;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 RKTNGFRVPL 12
 DB 197 RKTNGFRVPL 206

RESULT 13

Q912G3 PRELIMINARY; PRT; 631 AA.
 ID Q912G3;
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, last annotation update)
 DE Hypothetical protein PA1941.
 GN PA1941.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OK NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.";
 RL Nature 406:959-964 (2000).
 DR EMBL; AB004620; AAG05329.1; -.
 DR InterPro; IPR000345; CYC heme bind.
 DR PROSITE; PS00190; CYTOCHROME C; 2.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 631 AA; 70904 MW; E72A4632AEBB75F1 CRC64;

Query Match 50.0%; Score 41; DB 16; Length 631;
 Best Local Similarity 58.3%; Pred. No. 90;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 QKTNGFRVPL 13
 DB 474 QKTNGFRVPL 485

RESULT 14
 ID Q64483 PRELIMINARY; PRT; 876 AA.
 AC Q64483;
 GN

DT 01-AUG-1998 (TRENBLrel. 07, Created)
 DT 01-AUG-1998 (TRENBLrel. 07, last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, last annotation update)
 DE Putative receptor-like protein kinase (Senescence-induced receptor-
 like serine/threonine kinase).
 GN AT2G19190 OR SIRK.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
 OK NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kail S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buehl C.R., Ketchum K.A., Lee J.U., Rensing C.M., Koo H., Motilal K.S.,
 RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhagen G.P., Preuss D., Nieman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:761-768 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RA Robatzek S., Somasich I.E.;
 RT "Targets of AtWRKY6 regulation during plant senescence and pathogen
 defense.";
 RL Genes Dev. 0:0-0 (2002).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 DR EMBL; AC002392; AAD12037.1; -.
 DR EMBL; AF486619; AAL92103.1; -.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007090; LRR plant.
 DR InterPro; IPR000719; Prot_Kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; PF00560; LRR; 3.
 DR Pfam; PF00069; kinase; 1.
 DR ProDom; PD000001; Prot_Kinase; 1.
 DR PROSITE; PS05052; LRR_PS; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
 KW ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;
 Transferrase.
 SQ SEQUENCE 876 AA; 98711 MW; EEB1C3E335C7B6 CRC64;

Query Match 50.0%; Score 41; DB 10; Length 876;
 Best Local Similarity 46.2%; Pred. No. 13e+02;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

RESULT 15

Q8NT25 PRELIMINARY; PRT; 1333 AA.
 ID Q8NT25;
 AC Q8NT25;
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, last annotation update)
 DE DNA-directed RNA polymerase beta' subunit/160 kDa subunit (split gene
 in archaea and euk) (EC 2.7.7.6).
 GN GCL0489.

OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCTB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP005275; BAB97882.1; -;
DR InterPro: IPR000722; RNA_pol_A.
DR InterPro: IPR007080; RNA_pol_Rpb1_1.
DR InterPro: IPR007083; RNA_pol_Rpb1_3.
DR InterPro: IPR007083; RNA_pol_Rpb1_4.
DR InterPro: IPR007081; RNA_pol_Rpb1_5.
DR Pfam: PF04997; RNA_pol_Rpb1_1; 1.
DR Pfam: PF00623; RNA_pol_Rpb1_2; 1.
DR Pfam: PF04983; RNA_pol_Rpb1_3; 1.
DR Pfam: PF05000; RNA_pol_Rpb1_4; 1.
DR Pfam: PF04998; RNA_pol_Rpb1_5; 1.
KW DNA-directed RNA polymerase; Transferase; Nucleotidyltransferase;
KW Complete proteome.
SQ SEQUENCE 1333 AA; 147294 MW; 0F6B89ADCC719CF9 CRC64;

Query Match 48.8%; Score 40; DB 16; Length 1333;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 TRNGFRVPLARE 16
DB 903 TRQGVRFVPAAE 914
|||:|
|||:|

Search completed: July 30, 2003, 16:29:27
Job time : 29.1429 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 / Search time 5.2381 Seconds
(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-7

Sequence: 1 TORXTRNGFRVPLARE 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	712	1 PERL_BOVIN	P80025 bos taurus
2	59	72.0	712	1 PERL_HUMAN	P22079 homo sapien
3	51	62.2	718	1 PERM_MOUSE	P11247 mus musculu
4	46	56.1	716	1 PERM_MOUSE	P49290 mus musculu
5	44	53.7	421	1 ARGB_XYLFA	O9pemt xylalla fas
6	42	51.2	619	1 DNK_MYCL	P19993 mycobacteri
7	42	51.2	715	1 PERE_HUMAN	P11678 homo sapien
8	41	50.0	488	1 HP11_YEAST	Q12060 saccharomyc
9	41	50.0	876	1 SIRK_ARATH	O64483 arabidopsis
10	39	47.6	320	1 BXD1_DROME	P09276 drosophila
11	38	46.3	483	1 VPI9_VZVD	P44622 haemophilus
12	38	46.3	491	1 ILVC_HAEIN	O9c1f1 pasteurella
13	38	46.3	491	1 ILVC_PASWU	O9c1f1 pasteurella
14	38	46.3	897	1 SYI_FERPE	O9x442 feridobact
15	38	46.3	1066	1 HGPC_HAEIN	O9x442 feridobact
16	38	46.3	1067	1 HGBC_HAEIN	O9x442 feridobact
17	38	46.3	1084	1 HGBC_HAEIN	O9x442 feridobact
18	38	46.3	4594	1 DHC_HUMAN	P44836 haemophilus
19	38	46.3	4644	1 DHC_HUMAN	O14204 homo sapien
20	38	46.3	4644	1 DHC_MOUSE	O9jnu4 mus musculu
21	37	45.1	242	1 Y503_VIBVU	P38650 rattus norv
22	37	45.1	252	1 HIS6_RHOSH	O8del1 vibrio vuln
23	37	45.1	363	1 AG22_HUMAN	P50052 homo sapien
24	37	45.1	363	1 AG22_HUMAN	O92026 meriones un
25	37	45.1	363	1 AG22_MOUSE	P35374 mus musculu
26	37	45.1	363	1 AG22_MOUSE	P35374 mus musculu
27	37	45.1	465	1 INXB_RAT	O23157 rattus norv
28	37	45.1	642	1 NOGI_SCHPO	O94659 scitizosach
29	37	45.1	646	1 UVAB_METTH	O26462 methanobact
30	37	45.1	745	1 PERM_HUMAN	P05164 homo sapien
31	37	45.1	1014	1 ACAR_ARATH	O81108 arabidopsis
32	37	45.1	3027	1 POLG_PYFV1	O05057 parainfl yel
33	36	43.9	44	1 RL34_BRUME	O8y4a1 bruceella me

34	36	43.9	44	1 RL34_RALSO	O8y3h9 ralistonia s
35	36	43.9	44	1 RL34_RHILQ	O98d90 rhicobium l
36	36	43.9	242	1 Y538_VIBPA	O87d86 vibrio para
37	36	43.9	251	1 HIS6_LISIN	O92d88 listeria in
38	36	43.9	251	1 HIS6_LISMO	O9y9g5 listeria mo
39	36	43.9	255	1 HIS6_NEIMA	O9jvhs neisseria m
40	36	43.9	255	1 HIS6_NEIME	O9k0h4 neisseria m
41	36	43.9	257	1 UNG_DICDI	P53766 dictyostell
42	36	43.9	271	1 HIS6_ARCFU	O29439 archaeglob
43	36	43.9	446	1 TRME_CAUCR	O9x4f9 caulobacter
44	36	43.9	471	1 YIR1_DROME	P16424 drosophila
45	36	43.9	556	1 SR54_CANAL	O42816 candida alb

ALIGNMENTS

RESULT 1	ID	PERL_BOVIN	STANDARD	PRT	712 AA.
AC	P80025				
DT	01-AUG-1991 (Rel. 19, Created)				
DT	01-AUG-1991 (Rel. 19, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Lactoperoxidase precursor (EC 1.11.1.7) (LPO).				
GN	LPO.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_Taxid=9913;				
RN	[1]				
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.				
RX	MEDLINE=91025552; PubMed=2222811;				
RA	Dull T.J., Uyeda C., Stroberg A.D., Nedwin G., Seilhammer J.J.;				
RT	"Molecular cloning of cDNAs encoding bovine and human				
RT	lactoperoxidase.";				
RL	DNA Cell Biol. 9:499-509(1990).				
RN	[2]				
RP	SEQUENCE OF 101-712.				
RC	TISSUE=Milk;				
RX	MEDLINE=91266958; PubMed=2050150;				
RA	Cals M.-M., Maillart P., Brignon G., Anglade P., Ribadeau-Dumas B.;				
RT	"Primary structure of bovine lactoperoxidase, a fourth member of a				
RT	mammalian heme peroxidase family.";				
RL	Eur. J. Biochem. 198:733-739(1991).				
RN	[3]				
RP	CIRCULAR DICHOISM ANALYSIS, AND PARTIAL SEQUENCE.				
RX	MEDLINE=20384171; PubMed=10924350;				
RA	Watanabe S., Murata S., Kumura H., Nakamura S., Bollen A.;				
RT	"Bovine lactoperoxidase and its recombinant: comparison of structure				
RT	and some biochemical properties.";				
RL	Biochem. Biophys. Res. Commun. 274:756-761(2000).				
RN	[4]				
RP	COVALENT HEME ATTACHMENT.				
RX	MEDLINE=97238604; PubMed=9083001;				
RA	DePillis G.D., Ozaki S.-I., Kuo J.M., Maltby D.A.,				
RT	Ortiz de Montellano P.R.;				
RT	"Autocatalytic processing of heme by lactoperoxidase produces the				
RT	native protein-bound prosthetic group.";				
RL	J. Biol. Chem. 272:8857-8860(1997).				
RN	[5]				
RP	COVALENT HEME ATTACHMENT.				
RC	TISSUE=Milk;				
RX	MEDLINE=98447642; PubMed=9774411;				
RA	Rae T.D., Goff H.M.;				
RT	"The heme prosthetic group of lactoperoxidase. Structural				
RT	characteristics of heme 1 and heme 1-peptides.";				
RL	J. Biol. Chem. 273:27968-27977(1998).				
RN	[6]				
RP	COVALENT HEME ATTACHMENT, AND MUTAGENESIS OF ASP-225 AND GLU-375.				
RX	MEDLINE=21481384; PubMed=11597409;				

RA Suriano G., Matanabe S., Ghibaudo E.M., Bollen A., Ferrari R.P.,
 RA Moguljevsky N.;
 RT "Glut155Gln and Asp225Val mutants: about the nature of the covalent
 RT linkages between heme group and apo-protein in bovine
 RT Lactoperoxidase";
 RT Bioorg. Med. Chem. Lett. 11:2827-2831(2001).
 RN [7]
 RN COVALENT HEME ATTACHMENT, AND MUTAGENESIS OF ASP-225 AND GLU-375.
 RX MEDLINE=21850707; PubMed=11756449;
 RA Colas C., Kuo J.M., Ortiz de Montellano P.R.;
 RT "Asp-225 and Glu-375 in autocatalytic attachment of the prosthetic
 RT heme group of lactoperoxidase";
 RT J. Biol. Chem. 277:7191-7200(2002).
 CC -1- FUNCTION: LPO IS AN ANTIMICROBIAL AGENT. IT IS THOUGHT TO HELP
 CC PROTECT THE UDDER FROM INFECTION AND PROMOTE GROWTH IN NEBORN
 CC CALVES.
 CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
 CC -1- COFACTOR: Binds 1 protoheme IX and 1 iron(III) ion.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.
 CC -1- DATABASE: NAME=worthington-enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/1/LPO.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M58150; AAA62714.1; -
 CC PIR; A35828; A35828.
 DR HSSP; P05164; 1CXP.
 DR InterPro; IPR002007; Anim_peroxidase.
 DR InterPro; IPR002016; Peroxidase.
 DR Pfam; PF03098; An_peroxidase; 1.
 DR PRINTS; PR00457; ANPEROXIDASE.
 DR PROSITE; PS00435; PEROXIDASE_1; FALSE_NEG.
 DR PROSITE; PS00436; PEROXIDASE_2; FALSE_NEG.
 DR PROSITE; PS50292; PEROXIDASE_3; 1.
 DR PROSITE; PS50292; PEROXIDASE; Iron; Heme; Glycoprotein; Milk; signal;
 KW Antidiuretic.
 KW Antidiuretic.
 FT SIGNAL 1 22
 FT PROPEP 23 100
 FT CHAIN 101 712
 FT ACT_SITE 226 226
 FT ACT_SITE 372 372
 FT BINDING 225 225
 FT BINDING 375 375
 FT METAL 468 468
 FT CARBOHYD 106 106
 FT CARBOHYD 212 212
 FT CARBOHYD 322 322
 FT CARBOHYD 358 358
 FT CARBOHYD 449 449
 FT CARBOHYD 449 449
 FT VARIANT 394 394
 FT CONFLICT 449 449
 FT SEQUENCE 712 AA; 80642 MW; 28EBD4C0C8420BD CRC64;
 N -> R (IN REF. 2).
 Query Match 100.0%; Score 82; DB 1; Length 712;
 Best Local Similarity 100.0%; Pred. No. 8.5e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AC P22079; Q13408;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Lactoperoxidase precursor (EC 1.11.1.7) (LPO) (Salivary peroxidase)
 DE (SPO).
 GN LPO.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Submandibular gland;
 RC MEDLINE=97082979; PubMed=8964511;
 RA Kiser C., Caterina J., Engler J.A., Rahemtulla B., Rahemtulla F.;
 RT "Cloning and sequence analysis of the human salivary peroxidase-
 RT encoding cDNA";
 RT Gene 173:261-264(1996).
 RL [2]
 RN SEQUENCE OF 389-712 FROM N.A.
 RX MEDLINE=91025552; PubMed=2222811;
 RA Dull T.J., Uyeda C., Strosberg A.D., Nedwin G., Seilhamer J.J.;
 RT "Molecular cloning of cDNAs encoding bovine and human
 RT lactoperoxidase";
 RT DNA Cell Biol. 9:499-509(1990).
 CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
 CC -1- COFACTOR: Binds 1 protoheme IX and 1 iron(III) ion.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK AND SALIVARY GLAND.
 CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U9573; AAC50717.1; -
 CC EMBL; M58151; AAA63213.1; -
 DR PIR; JC4935; JC4935.
 DR HSSP; P05164; 1CXP.
 DR Genew; HGNC:6678; LPO.
 DR MIM; 150205; -
 DR GO; GO:0016686; P:lactoperoxidase activity; NAS.
 DR GO; GO:0006804; P:peroxidase reaction; NAS.
 DR InterPro; IPR002007; Anim_peroxidase.
 DR InterPro; IPR002016; Peroxidase.
 DR Pfam; PF03098; An_peroxidase; 1.
 DR PRINTS; PR00457; ANPEROXIDASE.
 DR PROSITE; PS00435; PEROXIDASE_1; FALSE_NEG.
 DR PROSITE; PS00436; PEROXIDASE_2; FALSE_NEG.
 DR PROSITE; PS50292; PEROXIDASE_3; 1.
 DR PROSITE; PS50292; PEROXIDASE; Iron; Heme; Glycoprotein; Milk; signal;
 KW Oxidoreductase; Peroxidase; Iron; Heme; Glycoprotein; Milk; signal;
 FT SIGNAL 1 26
 FT PROPEP 27 100
 FT CHAIN 101 712
 FT ACT_SITE 226 226
 FT ACT_SITE 372 372
 FT BINDING 225 225
 FT BINDING 375 375
 FT METAL 468 468
 FT CARBOHYD 106 106
 FT CARBOHYD 212 212
 FT CARBOHYD 322 322
 FT CARBOHYD 358 358
 FT CONFLICT 421 421
 FT SEQUENCE 712 AA; 80287 MW; BC7AAA410997198C0 CRC64;
 V -> M (IN REF. 2).
 Query Match 72.0%; Score 59; DB 1; Length 712;
 Best Local Similarity 75.0%; Pred. No. 0.01;

Matches	12: Conservative	1; Mismatches	3; Indels	0; Gaps	0;
0y	1	TQKTRNGFRVPLARE	16		
Db	179	TPGKTRNGFRVPLARE	194		
	PERM MOUSE	STANDARD;	PRT;	718 AA.	
ID	PERM MOUSE	STANDARD;	PRT;	718 AA.	
AC	P11247;				
DT	01-JUL-1989 (Rel. 11, Created)				
DT	01-JUL-1989 (Rel. 11, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Myeloperoxidase precursor (EC 1.11.1.7) (Mpo).				
GN	MPO.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxId=10090;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C3H;				
RC	MEDLINE=89345187; PubMed=2548170;				
RA	Venturelli D., Shirat N., Gempert I., Bittenbender S.,				
RA	Rovera G.;				
RT	"Nucleotide sequence of cDNA for murine myeloperoxidase.";				
RL	Nucleic Acids Res. 17:5852-5852(1989).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90016884; PubMed=2552419;				
RA	Venturelli D., Bittenbender S., Rovera G.;				
RT	"Sequence of the murine myeloperoxidase (Mpo) gene.";				
RL	Nucleic Acids Res. 17:7987-7988(1989).				
CC	-1- FUNCTION: Part of the host defense system of polymorphonuclear				
CC	leukocytes. It is responsible for microbicidal activity against a				
CC	wide range of organisms. In the stimulated PMN, MPO catalyzes the				
CC	production of hypohalous acids, primarily hypochlorous acid in				
CC	physiologic situations, and other toxic intermediates that greatly				
CC	enhance PMN microbicidal activity (By similarity).				
CC	-1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.				
CC	-1- CATALYTIC ACTIVITY: Cl(-) + H(2)O(2) = HOCl + 2 H(2)O.				
CC	-1- COFACTOR: Binds 1 prothemin IX covalently, 1 iron(III) ion and 1				
CC	calcium ion per subunit (By similarity).				
CC	-1- SUBUNIT: Tetramer of two light chains and two heavy chains.				
CC	-1- SUBCELLULAR LOCATION: Lysosomal.				
CC	-1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation in				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.1db-sib.ch/announce/or_send_an_email_to_license@1db-sib.ch).				
CC	-----				
CC	EMBL; X15313; CA33373.1; -.				
DR	EMBL; X15378; CA333439.1; -.				
DR	PIR; S06068; S06068.				
DR	HSP; P05164; ICRP.				
DR	MGI; MGI:97143; Mpo.				
DR	InterPro; IPR002007; Anim_peroxidase.				
DR	InterPro; IPR002016; Peroxidase.				
DR	PIR; PF03098; An_peroxidase; 1.				
DR	PRINTS; PR00457; ANPEROXIDASE.				
DR	PROSITE; PS00435; PEROXIDASE_1; 1.				
DR	PROSITE; PS00436; PEROXIDASE_2; FALSE_NEG.				
DR	PROSITE; PS02922; PEROXIDASE_3; 1.				
KW	Oxidoreductase; Peroxidase; Iron; Heme; Calcium-binding; Glycoprotein;				
KW	Signal; Lysosome.				
FT	SIGNAL	1	15		
FT	PROPEP	16	138	POTENTIAL.	
FT	CHAIN	139	718	MYELOPEROXIDASE.	

```

FT CHAIN 139 252 MYELOPEROXIDASE LIGHT CHAIN.
FT ACT_SITE 235 235 DISTAL HISTIDINE (POTENTIAL).
FT ACT_SITE 379 379 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT METAL 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 711 711 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 718 AA; 81070 MW; D734401FE8B4011B CRC64;

Query Match 62.2%; Score 51; DB 1; Length 718;
Best Local Similarity 81.8%; Pred. No. 0.27;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy 6 RNFGRVPLARE 16
Db 193 RNFGRVPLARQ 203

RESULT 4
PERE MOUSE
ID PERE MOUSE STANDARD; PRT; 716 AA.
AC P49290; Q61798;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Eosinophil peroxidase precursor (EC 1.11.1.7) (EPO).
GN EPX OR EPER.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Bone marrow;
RX MEDLINE=96369651; PubMed=8773591;
RA Horton M.A., Larson K.A., Lee J.J., Lee N.A.;
RT "Cloning of the murine eosinophil peroxidase gene (mEPO):
RT characterization of a conserved subgroup of mammalian hematopoietic
RT peroxidases."
RL J. Leukoc. Biol. 60:285-294(1996).
CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
CC -1- COFACTOR: Binds 1 prothemin IX and 1 Iron(III) ion.
CC -1- SUBUNIT: TRIMER OF TWO LIGHT CHAINS AND TWO HEAVY CHAINS (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF EOSINOPHILS.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration --
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D78353; BAAL1370.1; -;
DR EMBL; L77979; BAB40403.1; ALT_INIT.
DR HSSP; P05164; ICXP.
DR MGD; MG1:107569; Epx.
DR InterPro; IPR002007; Anim_peroxidase.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF03098; An_peroxidase; 1.
DR PRINTS; PRO0457; ANPEROXIDASE.
DR PROSITE; PS00435; PEROXIDASE 1; 1.

```


RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Frazer A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RA "Massive gene decay in the leprosy bacillus";
RA Nature 409:1007-1011(2001).
[3]
RN SEQUENCE OF 276-619 FROM N.A.
RP MEDLINE=89079288; PubMed=2491836;
RX Garbisa R.J., Hellqvist L., Booth R.J., Radford A.J., Britton W.J.,
RA Asbury L., Trent R.J., Basten A.;
RT "Homology of the 70-kilodalton antigens from Mycobacterium leprae and
RT Mycobacterium bovis with the Mycobacterium tuberculosis 71-kilodalton
RT antigen and with the conserved heat shock protein 70 of eucaryotes";
RL Infect. Immun. 57:204-212(1989).
[4]
RN PHOSPHORYLATION.
RP MEDLINE=98422335; PubMed=9748655;
RX Peake P., Winter N., Britton W.;
RA "Phosphorylation of Mycobacterium leprae heat-shock 70 protein at
RT threonine 175 alters its substrate binding characteristics.";
RL Biochim. Biophys. Acta 1387:387-394(1998).
CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC -1- INDUCTION: BY stress conditions e.g. heat shock (BY similarity).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M95576; AAA25362.1; -
DR EMBL; AL583925; GAC32013.1; -
DR PIR; B87221; B87221.
DR HSP; P08109; ICKR.
DR Phosbite; P19993; -
DR Leproma; M2496; -
DR HAMAP; MF_00332; -1.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PRINTS; PR00301; HEATSHOCK70.
DR ProDom; PD000089; Hsp70; 1.
DR PROSITE; PS00287; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
DR PROSITE; PS01036; HSP70_3; 1.
DR Chaperone; ATP-binding; Heat shock; Phosphorylation;
DR Complete proteome.
DR INIT MET 0
DR MOD RES 174 0
DR CONFLICT 369 174
DR CONFLICT 607 618
DR FT VDAEYVDDERES -> LTRKMTTNSGP (IN REF. 1
AND 3).
SQ SEQUENCE 619 AA; 66547 MW; 9490800A23A96612 CRC64;
Query Match 51.2%; Score 42; DB 1; Length 619;
Best Local Similarity 72.7%; Pred. No. 9.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

AC P11678;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Eosinophil peroxidase precursor (EC 1.11.1.7) (EPO).
GN EPX OR EPER OR EPO OR EPP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Placenta;
RX MEDLINE=89380315; PubMed=2550461;
RA Sakamaki K., Tomonaga M., Teukui K., Nagata S.;
RT "Molecular cloning and characterization of a chromosomal gene for
RT human eosinophil peroxidase";
RL J. Biol. Chem. 264:16828-16836(1989).
RN [2]
RP SEQUENCE OF 13-715 FROM N.A., AND SEQUENCE OF 140-187 AND 251-288.
RP TISSUE=Blood;
RX MEDLINE=89235593; PubMed=2541222;
RA Ten R.M., Pease L.R., McKean D.J., Bell M.P., Gleich G.J.;
RT "Molecular cloning of the human eosinophil peroxidase. Evidence for
RT the existence of a peroxidase multigene family";
RL J. Exp. Med. 169:1757-1769(1989).
RN [3]
RP COVALENT HEME ATTACHMENT, AND PARTIAL SEQUENCE.
RP TISSUE=Blood;
RX MEDLINE=9287891; PubMed=10358043;
RA Oxyvig C., Thomsen A.R., Overgaard M.T., Sorensen E.S., Hoejrup P.,
RA Bjerrum M.J., Gleich G.J., Sottrup-Jensen L.;
RT "Biochemical evidence for heme linkage through esters with Asp-93 and
RT Glu-241 in human eosinophil peroxidase. The ester with Asp-93 is only
RT partially formed in vivo";
RL J. Biol. Chem. 274:16953-16958(1999).
RN [4]
RP VARIANT EPD HIS-286.
RP MEDLINE=95107984; PubMed=7809065;
RA Romano M., Patriarca P., Melo C., Baralle F.E., Dri P.;
RT "Hereditary eosinophil peroxidase deficiency: immunochemical and
RT spectroscopic studies and evidence for a compound heterozygosity of
RT the defect";
RL Proc. Natl. Acad. Sci. U.S.A. 91:12496-12500(1994).
CC -1- CATALYTIC ACTIVITY: Donor + H(2)O(2) = oxidized donor + 2 H(2)O.
CC -1- COFACOR: Binds 1 protonome IX and 1 iron(III) ion.
CC -1- SUBUNIT: Tetramer of two light chains and two heavy chains.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES OF EOSINOPHILS.
CC -1- DISEASE: Defects in EPX are the cause of eosinophil peroxidase
CC deficiency (EPD) [MIM:261500]. EPD is an autosomal recessive
CC defect where anomalous eosinophils are characterized by nuclear
CC hypersegmentation, hypogranulation, and negative peroxidase and
CC phospholipid staining.
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. XPO SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M29913; AAA58458.1; -
DR EMBL; M29904; AAA58458.1; JOINED.
DR EMBL; M29905; AAA58458.1; JOINED.
DR EMBL; M29906; AAA58458.1; JOINED.
DR EMBL; M29907; AAA58458.1; JOINED.
DR EMBL; M29908; AAA58458.1; JOINED.
DR EMBL; M29909; AAA58458.1; JOINED.
DR EMBL; M29910; AAA58458.1; JOINED.
DR EMBL; M29911; AAA58458.1; JOINED.
DR EMBL; M29912; AAA58458.1; JOINED.

EMBL: X14346; CAA32530.1; -.
 DR PIR: A34408; A34408.
 DR HSSP: POS164; 1CXP.
 DR Gene: HGNC:3423; EPX.
 DR MIM: 131399; -.
 DR MIM: 261500; -.
 DR InterPro: IPR002007; Anim_peroxidase.
 DR InterPro: IPR002016; Peroxidase.
 DR Pfam: PF03098; An_peroxidase; 1.
 DR PRINTS: PR00457; ANPEROXIDASE.
 DR PROSITE: PS00435; PEROXIDASE_1; 1.
 DR PROSITE: PS00436; PEROXIDASE_2; FALSE_NEG.
 DR PROSITE: PS00292; PEROXIDASE_3; 1.
 DR Oxidoreductase; Peroxidase; Iron; Heme; Glycoprotein; Signal;
 KM Disease mutation.
 FT SIGNAL 1 17
 FT PROPEP 18 139
 FT CHAIN 140 250
 FT CHAIN 251 715
 FT BINDING 232 232
 FT BINDING 380 380
 FT ACT SITE 233 233
 FT ACT SITE 377 377
 FT METAL 474 474
 FT CARBOHYD 52 52
 FT CARBOHYD 113 113
 FT CARBOHYD 327 327
 FT CARBOHYD 363 363
 FT CARBOHYD 700 700
 FT CARBOHYD 708 708
 FT VARIANT 286 286
 FT CONFLICT 13 18
 FT CONFLICT 21 21
 FT CONFLICT 113 113
 FT CONFLICT 163 163
 FT CONFLICT 645 660
 SO SEQUENCE 715 AA; 81040 MW; CEBAB689A6C46374 CRG4;
 Query Match 51.2%; Score 42; DB 1; Length 715;
 Best Local Similarity 53.3%; Pred. No. 11;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 TORKTNGPRVPLAR 15
 DB 186 TFSRRRNGFLPLVR 200
 RESULT 8
 ID HFT1_YEAST STANDARD; PRT; 488 AA.
 AC Q12060; Q00039; Q02813;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transcriptional coactivator HFT1/ADA1.
 GN HFT1 OR ADA1 OR SUP10 OR YPL254W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxId=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BWGI-7A;
 RA MEDLINE=9729672; PubMed=9154821;
 RA Horichi J., Silverman N., Pina B., Marcus G.A., Guarente L.;
 RT "ADA1, a novel component of the ADA/GCN5 complex, has broader effects
 RT than GCN5, ADA2, or ADA3.";
 RL Mol. Cell. Biol. 17:3220-3228(1997).
 RL [2]
 RP SEQUENCE FROM N.A.

RA Brown N.G.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313271; PubMed=9169875;
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansgore W.,
 RA Araujo R., Aparicio A., Barrell B.G., Baccocck K., Bens V.,
 RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 RA Chung E., Churcher C.W., Coster F., Davis K., Davis R.W.,
 RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Dueserhoef A.,
 RA Duncan M., Fleeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 RA Hall J., Hebling U., Heumann K., Kaiman S., Kleine K.,
 RA Hunkeler-Smith S., Hyman R., Johnston M., Kaiman S., Kleine K.,
 RA Komp C., Kurd O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
 RA Marthe R., Messenguy F., Mewes H.-W., Mitiapati S., Moestl D.,
 RA Mueller-Auer S., Natch A., Nentwich U., Oetner P., Pearson D.,
 RA Petel F.X., Pohl T.M., Furnelle D., Schater M., Scharte M.,
 RA Scherene B., Schramm S., Schroeder M., Sidcu A.M., Tetelein H.,
 RA Ureastarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
 RA Walsh S.V., Wamburt R., Wang Y., Wedler E., Wedler H., Winnett E.,
 RA Walsh W.W., Zollner A., Vo D.H., Hani J.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
 RL Nature 387:103-105(1997).
 CC -1- FUNCTION: HFT1/ADA1 AND SPT20/ADAS MAY RECRUIT TATA BINDING
 CC PROTEIN (TBP) AND POSSIBLY OTHER BASAL FACTORS TO BIND TO THE TATA
 CC BOX WHILE ADA2, ADA3 AND GCN5 FUNCTION TO ACETYLATE NUCLEOSOMES
 CC OPENING UP THE PROMOTER REGION.
 CC -1- SUBUNIT: PART OF THE ADA/GCN5 COMPLEX THAT CONSISTS OF HFT1/ADA1,
 CC ADA2, ADA3, SPT20/ADAS AND GCN5.
 CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: U76735; AAB58359.1; -.
 CC EMBL: U41324; AAB4983.1; -.
 CC EMBL: Z67751; CAA91590.1; -.
 CC EMBL: Z73610; CAA97979.1; -.
 DR PIR: S61010; S61010.
 DR TRANSFAC: T04327; -.
 DR SGD: S0006175; HFT1.
 DR GO: GO:0000124; C:SAGA complex; IDA.
 DR GO: GO:0003712; P:transcription cofactor activity; IMP.
 DR GO: GO:0006366; P:transcription from Pol II promoter; IMP.
 KM Transcription regulation; Nuclear protein.
 FT CONFLICT 19 19
 FT CONFLICT 19 19
 SO SEQUENCE 488 AA; 54466 MW; B3ACF3C6101A541 CRG64;
 Query Match 50.0%; Score 41; DB 1; Length 488;
 Best Local Similarity 80.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 7 NGRVPLARE 16
 DB 253 NGRVPLARE 262
 RESULT 9
 ID SIRK_ARATH STANDARD; PRT; 876 AA.
 AC 064483;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Senescence-induced receptor-like serine/threonine kinase precursor
 DE (FLG22-induced receptor-like kinase 1).
 DR SIRK OR FRK1 OR AT2G19190 OR T20K24.21.
 GN

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21996065; PubMed=12000796;
 RA Robatzek S., Somssich I.E.;
 RT "Targets of AtWRKY6 regulation during plant senescence and pathogen
 defense.";
 RL Genes Dev. 16:1139-1149(2002).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Kounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Roming C.M., Koo H.L.,
 RA Moffat K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Unayam L.,
 RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Cressy T.H.,
 RA Goodman H.M., Somerville C.R., Copenhagen G.P., Preuss D.,
 RA Newman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:761-768(1999).
 RN [3]
 RP FUNCTION, AND INDUCTION.
 RX MEDLINE=21664699; PubMed=11875555;
 RA Asai T., Tena G., Plotnikova J., Willmann M.R., Chiu W.L.,
 RA Gomez-Gomez L., Bolter T., Ausubel F.M., Sheen J.;
 RT "MAP kinase signalling cascade in Arabidopsis innate immunity.";
 RL Nature 415:977-983(2002).
 CC -1- FUNCTION: Involved in innate immune response of plants.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- INDUCTION: Highly induced by WRKY22 or WRKY29 and by WRKY6 in
 CC senescent leaves. Also induced 30 minutes after flagellin
 CC treatment.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -1- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AA486619; AAL92103.1; -
 DR EMBL: AC002392; AAD12037.1; -
 DR PIR: T00540; T00540.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR007090; LRR plant.
 DR InterPro: IPR000719; LRR kinase.
 DR InterPro: IPR002290; Ser_Thr_kinase.
 DR Pfam: PF00560; LRR_3.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Prot_kinase; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KM Plant defense; Transferrase; Serine/threonine-protein kinase; Receptor;
 KM ATP-binding; Transmembrane; Repeat; Leucine-rich repeat; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 876
 FT DOMAIN 25 517
 FT TRANSMEM 518 538
 FT DOMAIN 539 876
 FT REPEAT 415 438
 FT LRR 1.

FT REPEAT 439 462 LRR 2.
 FT REPEAT 463 483 LRR 3.
 FT DOMAIN 574 847 PROTEIN KINASE.
 FT NP_BIND 580 588 ATP (BY SIMILARITY).
 FT ACT_SITE 687 697 BY SIMILARITY.
 FT BINDING 601 601 ATP (BY SIMILARITY).
 SQ SEQUENCE 876 AA; 98711 MW; EEE1C13E35BC76 CRC64;
 Query Match 50.0%; Score 41; DB 1; Length 876;
 Best Local Similarity 46.2%; Pred. No. 20;
 Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 TORXTRNGRVPVL 13
 Db 510 TKKRNKGVTIPL 522
 ID BOLD DROME STANDARD; PRT; 320 AA.
 AC 09VEB3; 095U84;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Brix domain containing protein 1 homolog.
 DE CG7993.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktarglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertan B.P., Bhattacharya D., Bolshakov S.,
 RA Borovda D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu L.B., Caceres A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.B., Gang N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan F., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh N.V., Mobarry C., Morris D., Moshrefi A.,
 RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,
 RA Palczolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Schebler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.F., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Weissman D.A., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).

```

RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brockstein P., Yu C., Champe M.,
  George R.A., Garin H., Krommiller B., Paclib J.M., Park S., Wan K.H.,
  Rubin G.M., Celniker S.E.;
  "A Drosophila full-length cDNA resource."
  J. Biol. Chem. 276:11581-11586 (2001).
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -1- SIMILARITY: Contains 1 Brix domain.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous
  gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A5003721; AAF5514.1; ALT_SEQ.
DR EMBL: AY058248; ALI3477.1; -.
DR Flybase: FBgn0038585; CG7993.
DR InterPro: IPR007109; Brix.
DR Pfam: PF04427; Brix. 1.
DR Hypothetical protein; Nuclear protein.
FT DOMAIN 30 243 Brix.
FT CONFICT 188 188 I -> T (IN REF. 1).
SQ SEQUENCE 320 AA; 36509 MW; EE98936DD68B3703 CRC64;

Query Match 47.6%; Score 39; DB 1; Length 320;
Best Local Similarity 69.2%; Pred. No. 16;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 KTRNGFRVPLAE 16
Db 10 KTRKGKVLARE 22

RESULT 11
VP19_VZVD STANDARD; PRT; 483 AA.
AC P09276;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Capsid assembly and DNA maturation protein (Capsid protein VP19C).
GN 20.
OS Varicella-zoster virus (strain Dumas) (VZV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirinae.
OX NCBI_TaxId=10338;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86306657; PubMed=3018124;
RA Davison A.J., Scott J.E.;
  "The complete DNA sequence of varicella-zoster virus."
  J. Gen. Virol. 67:1759-1816 (1986).
RL J. Gen. Virol. 67:1759-1816 (1986).
CC -1- FUNCTION: COMPONENT OF THE BASAL LAYER IN WHICH THE CAPSIDS ARE
  EMBEDDED. BINDS DNA (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP19C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04370; CAA27903.1; -.

```

```

DR PIR; B27343; WZB20.
DR InterPro: IPR004999; Herpes_VP19C.
DR Pfam; PF03327; Herpes_VP19C. 1.
DR Capsid assembly; Coat protein; DNA-binding.
SQ SEQUENCE 483 AA; 53971 MW; A584CF3D689F91 CRC64;

Query Match 46.3%; Score 38; DB 1; Length 483;
Best Local Similarity 57.1%; Pred. No. 37;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 TQKTRNGFRVPLA 14
Db 328 TQKQREGVRLHIA 341

RESULT 12
ILVC_HAEIN STANDARD; PRT; 491 AA.
AC P44822;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ketol-acid reductoisomerase (EC 1.1.1.86) (Acetohydroxy-acid
  isomeroeductase) (Alpha-keto-beta-hydroxyisocitil reductoisomerase).
GN ILVC OR H10682.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KW20 / ATCC 51907;
MEDLINE=9550630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
  Kiehlmann R.D., Lippman M.D., Martignetti J.L., McPhee J.D.,
  McKenney K., Sutton G., Shetty R., Liu L.-I., Glodek A., Kelley J.M.,
  Scott J.D., Shetty R., Liu L.-I., Glodek A., Kelley J.M.,
  Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
  Uetereback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
  Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
  Rasmussen C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
  Venter J.C.;
  "Whole-genome random sequencing and assembly of Haemophilus influenzae
  Rd."
  Science 269:496-512 (1995).
RL Science 269:496-512 (1995).
CC -1- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+)
  = (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
CC -1- PATHWAY: Valine and isoleucine biosynthesis; second step.
CC -1- SIMILARITY: Belongs to the ketol-acid reductoisomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32751; AAC22342.1; -.
DR PIR; B64086; B64086.
DR TIGR; H10682; -.
DR HAMAP; MF_00435; -. 1.
DR InterPro: IPR000506; ACh_Isomrctase.
DR Pfam; PF01450; ILVC; 1.
DR TIGRfam; TIGR00465; ILVC; 1.
KW Oxidoreductase; Branched-chain amino acid biosynthesis; NADP;
  Complete proteome.
FT INIT MET 0 BY SIMILARITY.
FT ACT SITE 131 131 POTENTIAL.
SQ SEQUENCE 491 AA; 54108 MW; 049C20AD56607076 CRC64;

Query Match 46.3%; Score 38; DB 1; Length 491;
Best Local Similarity 70.0%; Pred. No. 38;

```

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFRV 11
| | | | |
Db 79 QRTATNGFRV 88

RESULT 13
ID ILVC_PASMU STANDARD; PRT; 491 AA.
AC 09CFL1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ketol-acid reductoisomerase (EC 1.1.1.86) (Acetohydroxy-acid isomeroreductase) (Alpha-keto-beta-hydroxyisovaleryl reductoisomerase).
GN ILVC OR PM284;
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OX NCBI_Taxid=747;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.U., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.,
RT "Complete genomic sequence of Pasteurella multocida PM70.",
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -1- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+) = (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
CC -1- PATHWAY: Valine and isoleucine biosynthesis; second step.
CC -1- SIMILARITY: Belongs to the ketol-acid reductoisomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE006167; AAK03368.1; -.
DR HAMAP; MF_00435; -; 1.
DR InterPro; IPR000506; ACh_isomrdctse.
DR Pfam; PF01450; ILVC; 1.
DR TIGRfam; TIGR00465; ilvc; 1.
KW Oxidoreductase; Branched-chain amino acid biosynthesis; NADP; Complete proteome.
FT ACT SITE 132
SQ SEQUENCE 491 AA; 54204 MW; 9442457D56F2039B CRC64;

Query Match 46.3%; Score 38; DB 1; Length 491;
Best Local Similarity 70.0%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFRV 11
| | | | |
Db 80 QRTATNGFRV 88

RESULT 14
ID SYL_FERPB STANDARD; PRT; 697 AA.
AC 09XDB4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Isoleucyl-tRNA synthetase (EC 6.1.1.5) (Isoleucine--tRNA ligase) (Illes).
GN ILFS.
OS Feravidobacterium pennivorans.
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae;
OC Feravidobacterium.

OX NCBI_Taxid=93466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 9078;
RX MEDLINE=99240435; PubMed=10224005;
RA Bertoldo C., Duffner F., Jorgensen P.L., Antikarov G.,
RT "Pullulanase type I from feravidobacterium pennivorans: cloning, sequencing, and expression of the gene and biochemical characterization of the recombinant enzyme";
RL Appl. Environ. Microbiol. 65:2084-2091(1999).
CC -1- CATALYTIC ACTIVITY: ATP + L-isoleucine + tRNA(Ile) = AMP + diphosphate + L-isoleucyl-tRNA(Ile).
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF096862; AAD30388.1; -.
DR HSPB; P41972; IFPV.
DR InterPro; IPR002300; tRNA-synt_1a.
DR InterPro; IPR001412; tRNA-synt_1.
DR InterPro; IPR002301; tRNA-synt_1le.
DR Pfam; PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00984; TRNASYNTHLE.
DR TIGRfam; TIGR00392; ilcs; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding; Metal-binding; Zinc.
FT SITE 42 52 "HIGH" REGION.
FT SITE 575 579 "KMSKS" REGION.
FT BINDING 578 578 ATP (BY SIMILARITY).
SQ SEQUENCE 897 AA; 103511 MW; 729981A9C2278343 CRC64;

Query Match 46.3%; Score 38; DB 1; Length 897;
Best Local Similarity 58.3%; Pred. No. 69;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TORXTRNGFRV 12
| : | | | | |
Db 63 TRYKTRNGFRV 74

RESULT 15
ID HGPC_HAEIN STANDARD; PRT; 1066 AA.
AC 09X442;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin and hemoglobin-hapoglobin binding protein C precursor.
GN HGPC.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HI689 / Serotype B;
RX MEDLINE=99270928; PubMed=10338475;
RA Morton D.J., Whitty P.W., Stull T.L.;
RT "Effect of multiple mutations in the hemoglobin- and hemoglobin-hapoglobin-binding proteins, Hgpa, Hgpb, and Hgpc, of Haemophilus influenzae type B";
RL Infect. Immun. 67:2729-2739(1999).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE

```

CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPARIING.
CC ADDITION OR LOSS OF CCA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AF094574; AAD33112.1; -.
CC DR InterPro: IPR006970; PT.
CC DR InterPro: IPR000531; TonB_boxC.
CC DR Pfam: PF04886; PT; 1.
CC DR Pfam: PF00593; TonB_dep_Rec; 1.
CC DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
CC KM Outer membrane; Transport; TonB box; Multigene family; Signal;
CC Receptor; Repeat.
CC FT SIGNAL 1 24 POTENTIAL.
CC FT CHAIN 25 1066 HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN
CC FT DOMAIN 26 53 BINDING PROTEIN C.
CC FT REPEAT 26 29 7 X 4 AA TANDEM REPEATS OF Q-P-T-N.
CC FT REPEAT 30 33 1.
CC FT REPEAT 34 37 2.
CC FT REPEAT 38 41 3.
CC FT REPEAT 42 45 4.
CC FT REPEAT 46 49 5.
CC FT REPEAT 50 53 6.
CC FT SITE 50 53 7.
CC FT SITE 63 70 TONB BOX.
CC FT SITE 1049 1066 TONB C-TERMINAL BOX.
CC SQ SEQUENCE 1066 AA; 122593 MW; EF88BD5CE4247583 CRC64;

```

```

Query Match 46.3%; Score 38; DB 1; Length 1066;
Best Local Similarity 53.3%; Pred. NO. 82;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 2 QKRTNGFRVPLARE 16
DB 747 QKTYANGFRAPTSDE 761

```

Search completed: July 30, 2003, 16:24:32
Job time : 6.2381 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 10.1905 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-7
Perfect score: 82
Sequence: 1 TORKTRNGFRVPLARE 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 segs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	712	1 A35828	peroxidase (EC 1.1.1
2	59	72.0	712	2 UC4935	peroxidase (EC 1.1.1
3	51	62.2	718	2 S06068	myeloperoxidase (E
4	44	53.7	359	2 G82487	GODER family prote
5	44	53.7	438	2 H82734	acetylglutamate ki
6	42	51.2	344	2 A30544	dnak-type molecula
7	42	51.2	620	2 E87221	70 kd heat shock p
8	42	51.2	715	2 A34408	peroxidase (EC 1.1
9	41	50.0	138	2 C82758	hypothetical prote
10	41	50.0	414	2 T23940	hypothetical prote
11	41	50.0	488	2 S61010	hypothetical prote
12	41	50.0	631	2 B83404	H711 protein - yea
13	41	50.0	876	2 T00540	hypothetical prote
14	39	47.6	128	2 F71198	serine/threonine-s
15	39	47.6	235	2 B86442	hypothetical prote
16	39	47.6	461	2 T49693	probable 308 ribos
17	39	47.6	629	2 AD1838	hypothetical prote
18	39	47.6	684	2 G84730	Na+/H+ antiporter
19	38	46.3	123	2 AG1935	mutator-like trans
20	38	46.3	172	2 G70979	hypothetical prote
21	38	46.3	185	2 T36546	hypothetical prote
22	38	46.3	204	2 T46363	hypothetical prote
23	38	46.3	452	2 JC4100	hydroxylindole O-me
24	38	46.3	463	2 E70657	hydroxylindole O-me
25	38	46.3	483	1 WZB820	gene 20 protein -
26	38	46.3	492	1 B64086	ketol-acid reducto
27	38	46.3	529	2 T48253	myb-like protein -
28	38	46.3	628	2 A87596	hypothetical prote
29	38	46.3	641	2 T44585	acyl-CoA oxidase h

30	38	46.3	1084	2 B64086	hemoglobin-binding
31	38	46.3	1241	2 T18311	hypothetical prote
32	38	46.3	4644	1 A38905	dynein heavy chain
33	38	46.3	6420	2 T30283	polypeptide synthas
34	37	45.1	123	2 E72680	hypothetical prote
35	37	45.1	128	2 AB1464	hypothetical prote
36	37	45.1	146	2 G90560	B. subtilis Divic
37	37	45.1	172	2 R87000	hypothetical prote
38	37	45.1	181	2 H87384	probable membrane
39	37	45.1	252	2 S54838	acetyltransferase,
40	37	45.1	345	2 T28213	cyclase hlsF - Rho
41	37	45.1	363	2 T48261	DNA-directed RNA p
42	37	45.1	363	2 A49092	angiogenesis II rec
43	37	45.1	363	2 UC2543	angiogenesis II rec
44	37	45.1	375	2 T39329	hypothetical prote
45	37	45.1	417	2 T29770	hypothetical prote

ALIGNMENTS

RESULT 1
A35828
peroxidase (EC 1.1.1.7) precursor - bovine
M/Alternate names: lactoperoxidase
C/Species: Bos primigenius taurus (cattle)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Nov-2000
R/Accession: A35828, S16103
R/DUI, T.V./ Uyeda, C./ Strohsberg, A.D., Nedwin, G., Sellhammer, J.J.
DNA Cell Biol. 9, 499-509, 1990
A/Title: Molecular cloning of cDNAs encoding bovine and human lactoperoxidase.
A/Reference number: A35828, MUID:91025552, PMID:2222811
A/Accession: A35828
A/Molecule type: mRNA
A/Residues: 1-712 <DUI>
A/Cross-references: GB:58150; NID:9163306; PIDN:AA62714.1; PID:9163307
R/Cals, M.M./ Maillet, P.; Brignon, G.; Anglade, P.; Dumas, B.R.
Eur. J. Biochem. 198, 733-739, 1991
A/Title: Primary structure of bovine lactoperoxidase, a fourth member of a mammalian he
A/Reference number: S16103; MUID:9126958; PMID:2050150
A/Accession: S16103
A/Molecule type: protein
A/Residues: 101-108, 'B', '110-448, 'R', '450-712 <CAL>
C/Suprafamily: myeloperoxidase; myeloperoxidase homology
C/Key words: chromoprotein; glycoprotein; heme; iron; metalloprotein; oxidoreductase
F/1-21/Domain: signal sequence #status predicted <SIG>
F/32-100/Domain: signal sequence #status predicted <SIG>
F/101-712/Product: peroxidase #status predicted <PRO>
F/101-712/Domain: myeloperoxidase homology <MPX>
F/106-212, 322-358, 449/Binding site: carbohydrate (Asn) (covalent) #status predicted
F/132-145, 246-256, 280-274, 354-365, 573-630, 671-696/Disulfide bonds: #status predicted
F/226/Active site: His (distal axial ligand) #status predicted
F/372/Active site: Arg #status predicted
F/468/Binding site: heme/iron (His) (proximal axial ligand) #status predicted

Query Match
Best Local Similarity 100.0%; Score 82; DB 1; Length 712;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TORKTRNGFRVPLARE 16
Db 179 TORKTRNGFRVPLARE 194

RESULT 2
UC4935
peroxidase (EC 1.1.1.7) precursor - human
N/Alternate names: lactoperoxidase (LPO); salivary peroxidase (SPO)
C/Species: Homo sapiens (man)
C/Date: 22-Oct-1996 #sequence_revision 01-Nov-1996 #text_change 01-Dec-2000
R/Accession: UC4935, B35828
R/Kiser, C.; Caterina, J.; Engler, J.A.; Rahemtulla, B.; Rahemtulla, F.

```

F./234,362,383/Binding site: hemediol (asp, Glu, Met) (covalent) #status predicted
F./235/Active site: His (distal axial ligand) #status predicted
F./379/Active site: Arg (status predicted)
F./476/Binding site: hemediol iron (His) (proximal axial ligand) #status predicted

Query Match      62.2%; Score 51; DB 2; Length 718;
Best Local Similarity 81.8%; Pred. No. 0.48;
Matches          9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Oy              6 RRGFRVPLARE 16
                |||||:||||:
Db             193 RRGFRVPLARQ 203

RESULT 4
GGDEF family protein VCA0217 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
G.Species: Vibrio cholerae
C.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C.Accession: G82487
R.Heldberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Charston, D.; Emolaeva, M.D.; Yamahew, U.; Baas, S.; Qin, H.; Dragot, I.; Sellers, P.
I. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A.Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A.Reference number: AB2035; MID:20406833; PMID:10952301
A.Accession: G82487
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-359 <HEI>
A.Cross-references: GB:AE004361; GB:AE003853; MID:99657600; PIDN:AAP96129.1; GSPDB:GN001
A.Experimental source: serogroup O1; strain N16961; biotype RI Tor
A.Genetics:
A.Gene: VCA0217
A.Map position: 2

Query Match      53.7%; Score 44; DB 2; Length 359;
Best Local Similarity 69.2%; Pred. No. 4.3;
Matches          9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy              2 ORKTRNGFRVPLA 14
                |||||:||||
Db             74 OKTTPMGFRPLPA 86

RESULT 5
H82734
acetylglutamate kinase XF1001 [imported] - Xylella fastidiosa (strain 945c)
C.Species: Xylella fastidiosa
C.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C.Accession: H82734
R.anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A.Reference number: AB2515; MID:20365717; PMID:10910347
A>Note: for a complete list of authors see reference number A59328 below
A.Accession: H82734
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-438 <SIM>
A.Cross-references: GB:AE003938; GB:AE003849; NID:99105935; PIDN:AAF3811.1; GSPDB:GN001
A.Experimental source: strain 945c
R.Simpson, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carer, B.
Birones, M.R.S.; Docena, C.; El-Dorty, H.; Facincani, A.P.; Ferreira, A.J.S.
as-Neto, E.;
submitted to Genbank, June 2000
A.Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohn
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins,
A.Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
Rodrigues, V.; Rosa, A.U. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Savasat
F./Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
```

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XFI001

Query Match 53.7%; Score 44; DB 2; Length 438;
 Best Local Similarity 75.0%; Pred. No. 5.3;
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFRVPL 13
 DB 90 ORKTRNGFRVTL 101

RESULT 6
 A30544
 drak-type molecular chaperone - Mycobacterium leprae (fragment)
 N:Alternate names: 70K antigen
 C:Species: Mycobacterium leprae
 C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 13-Mar-1998
 C:Accession: A30544
 R:Garcia, R.J.; Hellqvist, L.; Booth, R.J.; Radford, A.J.; Britton, W.J.; Ashbury, L.; T
 Infect. Immun. 57, 204-212, 1989
 A:Title: Homology of the 70-kilodalton antigens from Mycobacterium leprae and Mycobacter
 0 of eucaryotes.
 A:Reference number: A30544; PMID:89079288; PMID:2491836
 A:Accession: A30544
 A:Molecule type: DNA
 A:Residues: 1-344 <GAR>
 A:Function:
 A:Description: Involved in protein folding and assembling/dissassembling of protein comp
 C:Superfamily: heat shock protein 70
 C:Keywords: ATP; molecular chaperone

Query Match 51.2%; Score 42; DB 2; Length 344;
 Best Local Similarity 72.7%; Pred. No. 9.5;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFRVPL 12
 DB 251 ORKTRNGSRVP 261

RESULT 7
 E87221
 70 kD heat shock protein (molecular chaperone) [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: E87221
 R:Cooley, S.T.; Eiglsmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
 R.; Davies, R.M.; Devlin, K.; Duchoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 eam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
 A:Title: Massive gene decay in the leprosy bacillus.
 A:Reference number: A86909; PMID:21128732; PMID:11234002
 A:Accession: E87221
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-620 <STO>
 A:Cross-references: GB:A1450380; NID:G13094043; PIDN:CAC32013.1; GSPDB:GMO0147
 C:Genetics:
 A:Gene: drak

C:Superfamily: heat shock protein 70

Query Match 51.2%; Score 42; DB 2; Length 620;
 Best Local Similarity 72.7%; Pred. No. 18;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFRVPL 12
 DB 527 ORKTRNGSRVP 537

RESULT 8
 A34408
 peroxidase (EC 1.11.1.7), eosinophil, precursor - human

C:Species: Homo sapiens (man)
 C:Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 24-Oct-2000
 C:Accession: A34408; B60866; S04746
 R:Sakamaki, K.; Tomonaga, M.; Tsubako, K.; Nagata, S.
 J. Biol. Chem. 264, 16828-16836, 1989

A:Title: Molecular cloning and characterization of a chromosomal gene for human eosinoph
 A:Reference number: A34408; PMID:89380315; PMID:2550461
 A:Accession: A34408
 A:Molecule type: DNA

A:Residues: 1-715 <SAR>
 A:Cross-references: GB:M29904; GB:M29905; GB:M29906; GB:M29907; GB:M29908; GB:M29909; GE
 R:Weller, P.F.; Ackerman, S.J.; Smith, J.A.
 J. Leukoc. Biol. 43, 1-4, 1988

A:Title: Eosinophil granule cationic proteins: major basic protein is distinct from the
 A:Reference number: A60866; PMID:88089268; PMID:3422083
 A:Accession: B60866

A:Molecule type: protein

A:Residues: 'XX', 142-150, 'XX', 153-154, 'EXX', 158-159 <WEL>
 A:Note: this sequence begins from the amino terminus of the light chain
 A:Ren, R.M.; Pease, L.R.; McKean, D.J.; Bell, M.P.; Gletch, G.J.
 J. Exp. Med. 169, 1757-1769, 1989

A:Title: Molecular cloning of the human eosinophil peroxidase. Evidence for the existen
 A:Reference number: S04746; PMID:89235593; PMID:2541222
 A:Accession: S04746

A:Molecule type: mRNA

A:Residues: 'EFGQD', 19-20, 'Q', 22-112, 'T', 114-644, 'ETETSGGRT', 655, 'C', 658, 'HQ', 661-715
 A:Cross-references: EMBL:X14346; NID:G31182; PIDN:CAA32530.1; PID:G31183
 A:Note: 163-Cys was also found

A:Note: the authors translated the codon ATT for residue 113 as Asn
 A:Note: part of this sequence, including the amino ends of both the heavy and light cha
 C:Genetics:
 A:Map position: 17

C:Superfamily: myeloperoxidase; myeloperoxidase homology

C:Keywords: chromoprotein; eosinophil; glycoprotein; heme; iron; metalloprotein; oxidore

F:36-712/Dominant: myeloperoxidase homology <MP>
 F:140-250/Product: peroxidase light chain #status experimental <LCH>
 F:251-715/Product: peroxidase heavy chain #status experimental <HCH>
 F:52,113,327,363,708/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:141-152,253-263,257-281,359-370,578-635,676-701/Disulfide bonds: #status predicted
 F:232,380/Binding site: heme/hemol (Asp, Glu) (covalent) #status predicted
 F:233/Active site: His (distal axial ligand) #status predicted
 F:377/Active site: Arg #status predicted
 F:474/Binding site: heme/hemol iron (His) (proximal axial ligand) #status predicted

Query Match 51.2%; Score 42; DB 2; Length 715;
 Best Local Similarity 53.3%; Pred. No. 20;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 TORKTRNGFRVPLAR 15
 DB 186 TFSRRNRGFLTLVLR 200

RESULT 9
 C82758
 hypothetical protein XF0830 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: C82758
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; PMID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: C82758
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-138 <SIM>

submitted to the Protein Sequence Database, May 1996
A:/Reference number: S64899
A:/Accession: S65283
A:/Molecule type: DNA
A:/Residues: 1-488 <POW>
A:/Cross-references: EMBL:Z73610, NID:gl370521, PIDN:CAA97979.1; PID:e246957; PID:gl37052
A:/Experimental source: strain S288C (AB972)
R:/Messenger: F.; Dubois, E.; Vierendeels, F.; Scherens, B.
submitted to the Protein Sequence Database, May 1996
A:/Reference number: S64935
A:/Accession: S65285
A:/Molecule type: DNA
A:/Residues: 1-111 <MES>
A:/Cross-references: EMBL:Z73610; MIPS:YPL254W
A:/Experimental source: strain S288C (AB972)
C:/Genetics:
A:/Gene: SGD:HFT1
A:/Cross-references: SGD:S0006175; MIPS:YPL254W
A:/Map position: 16L

Query Match	50.0%;	Score 41;	DB 2;	Length 488;
Best Local Similarity	80.0%;	Pred. No. 21;		
Matches 8;	Conservative	0;	Mismatches	2;
			Indels	0;
			Gaps	0;

QY 7 NGFRVPLARE 16
|||||

```

RESULT 12
B83404
hypothetical protein PA1941 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83404
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437357; PMID:10984043
A:Accession: B83404
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-631 <STO>
A:Cross-references: GB:AE004620; GB:AE004091; NID:G9947929; PIDN:AAG05329.1; GSPDB:GN001
C:Genetics:
A:Experimental source: strain PAO1
A:Gene: PA1941

Query Match          50.0%; Score 41; DB 2; Length 631;
Best Local Similarity 58.3%; Pred. No. 27;
Matches      7; Conservative    3; Mismatches      2; Indels      0; Gaps      0;

2 ORKTRNGFRVPL 13
|::|||::|:

```

RESULT 13
T00540
serine/threonine-specific protein kinase (EC 2.7.1.-) T20K24.21 - Arabidopsis thaliana
N|Alternate names: probable receptor-like protein kinase At2g19190
C|Species: Arabidopsis thaliana (mouse-ear cress)
C|Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001
C|Accession: T00540; F84573
R|Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, July 1997
A|Description: Arabidopsis thaliana chromosome II BAC T20K24 genomic sequence.
A|Reference number: Z14167
A|Accession: T00540
A|Status: translated from GB/EMBL/DBJ
M|Molecule type: DNA

P. 2

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using SW model

Run on: July 30, 2003, 16:23:41 ; Search time 11.2361 Seconds

(without alignments)
169.082 Million cell updates/sec

Title: US-09-787-070-7

Perfect score: 82
Sequence: 1 TORXTRNGFRVPLARE 16Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

```
1: Published Applications AA.*
2: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	48.8	264	US-10-156-761-12999	Sequence 12999, A
2	40	48.8	1333	US-09-738-626-4048	Sequence 4048, Ap
3	39	47.6	1156	US-09-757-716-1	Sequence 1, Appl1
4	39	47.6	1325	US-09-864-761-35612	Sequence 35612, A
5	38	46.3	124	US-09-984-271-247	Sequence 247, App
6	38	46.3	125	US-09-984-271-192	Sequence 192, App
7	38	46.3	172	US-10-080-170-556	Sequence 556, App
8	38	46.3	130	US-10-153-668-116	Sequence 116, App
9	38	46.3	492	US-09-815-242-11079	Sequence 11079, A
10	38	46.3	502	US-09-738-626-4722	Sequence 4722, Ap
11	37	45.1	15	US-10-133-056-10	Sequence 10, Appl
12	37	45.1	50	US-09-993-844-38	Sequence 38, Appl
13	37	45.1	172	US-10-080-170-74	Sequence 74, Appl
14	37	45.1	363	US-09-802-640-40	Sequence 40, Appl
15	37	45.1	363	US-10-133-056-2	Sequence 2, Appl1

16	37	45.1	363	US-10-133-056-4	Sequence 4, Appl1
17	37	45.1	305	US-10-225-567A-330	Sequence 330, App
18	37	45.1	403	US-09-738-626-6717	Sequence 6717, Ap
19	37	45.1	420	US-09-946-142-2	Sequence 2, Appl1
20	37	45.1	429	US-10-156-761-13577	Sequence 13577, A
21	36	43.9	221	US-10-080-170-46	Sequence 46, Appl
22	36	43.9	330	US-10-156-761-9078	Sequence 9078, Ap
23	35	42.7	48	US-09-925-299-1098	Sequence 1098, Ap
24	35	42.7	48	US-09-925-299-1098	Sequence 1098, Ap
25	35	42.7	84	US-09-764-891-3365	Sequence 3365, Ap
26	35	42.7	84	US-10-091-572-254	Sequence 254, App
27	35	42.7	107	US-09-864-761-42419	Sequence 42419, A
28	35	42.7	129	US-10-106-698-4874	Sequence 4874, Ap
29	35	42.7	135	US-09-828-366-26	Sequence 26, Appl
30	35	42.7	135	US-10-028-072-276	Sequence 276, App
31	35	42.7	135	US-10-121-049-276	Sequence 276, App
32	35	42.7	135	US-10-123-904-276	Sequence 276, App
33	35	42.7	135	US-10-140-470-276	Sequence 276, App
34	35	42.7	135	US-10-175-746-276	Sequence 276, App
35	35	42.7	135	US-10-176-918-276	Sequence 276, App
36	35	42.7	135	US-10-176-921-276	Sequence 276, App
37	35	42.7	135	US-10-137-865-276	Sequence 276, App
38	35	42.7	135	US-10-140-474-276	Sequence 276, App
39	35	42.7	135	US-10-142-431-276	Sequence 276, App
40	35	42.7	135	US-10-143-114-276	Sequence 276, App
41	35	42.7	135	US-10-140-002-276	Sequence 276, App
42	35	42.7	135	US-10-142-419-276	Sequence 276, App
43	35	42.7	135	US-10-123-262-276	Sequence 276, App
44	35	42.7	135	US-10-142-433-276	Sequence 276, App
45	35	42.7	135	US-10-121-050-276	Sequence 276, App

ALIGNMENTS

```
RESULT 1
US-10-156-761-12999
; Sequence 12999, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HOSIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12999
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12999
Query Match          48.8%; Score 40; DB 15; Length 264;
Best Local Similarity 64.3%; Pred. No. 31;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 ORKTRNGFRVPLAR 15
DB 144 ORKTRNGFRVPLAR 154
```

RESULT 2
US-09-738-626-4048

Sequence 4048, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: MAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIHO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, MAKOTO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASARU
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: JP 00/280988
NUMBER OF SEQ. ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 4048
LENGTH: 1333
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-4048

Query Match 48.8%; Score 40; DB 10; Length 1333;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 TRNGFRVPLARE 16
DB 903 TRQGVRRVVAAB 914

RESULT 3
US-09-757-716-1
Sequence 1, Application US/09757716
Patent No. US20010012515A1
GENERAL INFORMATION:
APPLICANT: Magna, Holly
APPLICANT: Schaffer, Paul
APPLICANT: Lawton, Michael
APPLICANT: Vocum, Sue
APPLICANT: Mitchell, Peter
APPLICANT: Hutchinson, Nancy
APPLICANT: Murty, Lynn E.
TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2
FILE REFERENCE: PF-0420 US
CURRENT APPLICATION NUMBER: US/09/757,716
CURRENT FILING DATE: 2001-01-09
PRIOR APPLICATION NUMBER: US/08/996,083
PRIOR FILING DATE: 1997-12-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1156
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte Clone No. US20010012515A1: 1388013
US-09-757-716-1

Query Match 47.6%; Score 39; DB 9; Length 1156;
Best Local Similarity 70.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 6 RNGFRVPLAR 15
DB 856 RRGFRINLAK 865

RESULT 4
US-09-864-761-35612
Sequence 35612, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecmtica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 35612
LENGTH: 1325
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000511.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
OTHER INFORMATION: EST HUMAN HIT: AW867076.1, EVALUATE 1.00e-103
OTHER INFORMATION: EST HUMAN HIT: AW867076.1, EVALUATE 1.00e-103

OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUATION: 5.00e-63
US-09-864-761-35612

Query Match 47.6%; Score 39; DB 9; Length 1325;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 TRNGFRVPLAR 16
DB 532 TRNGDRTPPLANE 543

RESULT 5

US-09-984-271-247
Sequence 247, Application US/09984271
Publication No. US20030040088A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 71 Human Secreted Proteins
FILE REFERENCE: P2030P1
CURRENT APPLICATION NUMBER: US/09/984,271
PRIOR FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: 09/482,273
PRIOR FILING DATE: 2000-01-13
PRIOR APPLICATION NUMBER: PCT/US99/15849
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: 60/092,921
PRIOR FILING DATE: 1998-07-15
PRIOR APPLICATION NUMBER: 60/092,922
PRIOR FILING DATE: 1998-07-15
PRIOR APPLICATION NUMBER: 60/092,956
PRIOR FILING DATE: 1998-07-15
NUMBER OF SEQ ID NOS: 267
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 247
LENGTH: 124
TYPE: PRT
ORGANISM: Homo sapiens
US-09-984-271-247

Query Match 46.3%; Score 38; DB 11; Length 124;
Best Local Similarity 77.8%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 KTRNGFRVP 12
DB 112 KTRNEFRPLP 120

RESULT 6

US-09-984-271-192
Sequence 192, Application US/09984271
Publication No. US20030040088A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 71 Human Secreted Proteins
FILE REFERENCE: P2030P1
CURRENT APPLICATION NUMBER: US/09/984,271
PRIOR FILING DATE: 2001-10-29
PRIOR APPLICATION NUMBER: 09/482,273
PRIOR FILING DATE: 2000-01-13
PRIOR APPLICATION NUMBER: PCT/US99/15849
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: 60/092,921
PRIOR FILING DATE: 1998-07-15
PRIOR APPLICATION NUMBER: 60/092,922
PRIOR FILING DATE: 1998-07-15
PRIOR APPLICATION NUMBER: 60/092,956
PRIOR FILING DATE: 1998-07-15
NUMBER OF SEQ ID NOS: 267
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 192
LENGTH: 125

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (125)
OTHER INFORMATION: Xaa equals stop translation
US-09-984-271-192

Query Match 46.3%; Score 38; DB 11; Length 125;
Best Local Similarity 77.8%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 KTRNGFRVP 12
DB 112 KTRNEFRPLP 120

RESULT 7

US-10-080-170-596
Sequence 596, Application US/10080170
Publication No. US20030129601A1
GENERAL INFORMATION:
APPLICANT: COLE, S.T.
TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
TREATMENT OF MYCOBACTERIOSES
FILE REFERENCE: 03495.0218
CURRENT APPLICATION NUMBER: US/10/080,170
PRIOR FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: 60/270,123
PRIOR FILING DATE: 2001-02-22
NUMBER OF SEQ ID NOS: 652
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 596
LENGTH: 172
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-10-080-170-596

Query Match 46.3%; Score 38; DB 16; Length 172;
Best Local Similarity 63.6%; Pred. No. 44;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 TRNGFRVPLAR 15
DB 102 TRSGIDIPPLAR 112

RESULT 8

US-10-153-668-116
Sequence 116, Application US/10153668
Publication No. US20030092616A1
GENERAL INFORMATION:
APPLICANT: HONDA, Goichi
APPLICANT: MATSUJIMA, Akio
APPLICANT: MURAMATSU, Shuji
APPLICANT: ISHIZAKA, Kenya
TITLE OF INVENTION: STAT6 Activating Gene
FILE REFERENCE: 1254-0207P
CURRENT APPLICATION NUMBER: US/10/153,668
PRIOR FILING DATE: 2002-05-24
PRIOR APPLICATION NUMBER: US 60/293,172
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: US 60/316,031
PRIOR FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: US 60/328,403
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: JP 2001-157043
PRIOR FILING DATE: 2001-05-25
PRIOR APPLICATION NUMBER: JP 2001-260681
PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: JP 2001-313175
PRIOR FILING DATE: 2001-10-10

NUMBER OF SEQ ID NOS: 488
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 116
 LENGTH: 330
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-153-668-116

Query Match 46.3%; Score 38; DB 15; Length 330;
 Best Local Similarity 87.5%; Pred. No. 88;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ORKTRNGFV 9
 |||||:
 Db 137 ORKTRNGFV 144

RESULT 9
 US-09-815-242-11079

Sequence 11079, Application US/09815242
 Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.

APPLICANT: Zykkind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

PRIOR FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 11079

LENGTH: 492

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-09-815-242-11079

Query Match 46.3%; Score 38; DB 9; Length 492;
 Best Local Similarity 70.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFV 11
 |||||:
 Db 80 ORKTRNGFV 89

RESULT 10

US-09-738-626-4722

Sequence 4722, Application US/09738626

Publication No. US20020197605A1

GENERAL INFORMATION:

APPLICANT: MAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO
 APPLICANT: HAYASHI, MIKIRO
 APPLICANT: OCHIAI, KEIKO
 APPLICANT: YOKOI, HARUHIKO
 APPLICANT: TATEISHI, NAKO
 APPLICANT: SENOH, AKIHIRO
 APPLICANT: IKEDA, MASATO
 APPLICANT: OZAKI, AKIO
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-125
 CURRENT APPLICATION NUMBER: US/09/738,626
 CURRENT FILING DATE: 2000-12-18
 PRIOR APPLICATION NUMBER: JP 99/377484
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: JP 00/159162
 PRIOR FILING DATE: 2000-04-07
 PRIOR APPLICATION NUMBER: JP 00/260988
 PRIOR FILING DATE: 2000-08-03
 NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: Patentin ver. 3.0
 SEQ ID NO 4722
 LENGTH: 502
 TYPE: PRT
 ORGANISM: Corynebacterium glutamicum
 US-09-738-626-4722

Query Match 46.3%; Score 38; DB 10; Length 502;
 Best Local Similarity 53.3%; Pred. No. 1.4e+02;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 ORKTRNGFVPLARE 16
 |||||:
 Db 454 ORKTRNGFVPLARE 468

RESULT 11

US-10-133-056-10

Sequence 10, Application US/10133056

Publication No. US2003008339A1

GENERAL INFORMATION:

APPLICANT: Tamura, Masaaki

TITLE OF INVENTION: Compositions and Methods of Treating Colorectal Polyps and Cancer

FILE REFERENCE: Vanderbilt Ref. No. US2003008339A1 V00112; Our Ref. No. US200300833

CURRENT APPLICATION NUMBER: US/10/133,056

CURRENT FILING DATE: 2002-04-26

NUMBER OF SEQ ID NOS: 12

SOFTWARE: Patentin version 3.1

SEQ ID NO 10

LENGTH: 15

TYPE: PRT

ORGANISM: Mus musculus

US-10-133-056-10

Query Match 45.1%; Score 37; DB 15; Length 15;
 Best Local Similarity 58.3%; Pred. No. 4.8;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ORKTRNGFVPL 13
 |||||:
 Db 1 ORKTRNGFVPL 12

RESULT 12

US-09-993-844-38

Sequence 38, Application US/09993844

Patent No. US20020106739A1

GENERAL INFORMATION:

APPLICANT: Oakley, Robert H.

APPLICANT: Barak, Lawrence S.

APPLICANT: Laporte, Stephanie A.

APPLICANT: Caron, Marc G.

TITLE OF INVENTION: Modified G-Protein Coupled Receptors

FILE REFERENCE: 033072-026

```

; CURRENT APPLICATION NUMBER: US/09/993,844
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: US 60/245,772
; PRIOR FILING DATE: 2000-11-03
; PRIOR APPLICATION NUMBER: US 60/260,363
; PRIOR FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 38
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-844-38

```

```

Query Match      45.1%; Score 37; DB 10; Length 50;
Best Local Similarity 58.3%; Pred. No. 17;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 ORKTRNGFRVPL 13
DB      13 OOKLRSVFRVPI 24

```

```

RESULT 13
US-10-080-170-74
; Sequence 74, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLB, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 74
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170-74

```

```

Query Match      45.1%; Score 37; DB 16; Length 172;
Best Local Similarity 63.6%; Pred. No. 66;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      5 TRNGFRVPLAR 15
DB      102 TRSGVDIPLAR 112

```

```

RESULT 14
US-09-802-640-40
; Sequence 40, Application US/09802640
; Publication No. US20030036057A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Andreas
; APPLICANT: Bonnal Aruna
; APPLICANT: Kiehn Patrick
; TITLE OF INVENTION: GENES AND POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE AND THEIR USE
; FILE REFERENCE: 24736-2048
; CURRENT APPLICATION NUMBER: US/09/802,640
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapien

```

```

US-09-802-640-40

```

```

Query Match      45.1%; Score 37; DB 11; Length 363;
Best Local Similarity 58.3%; Pred. No. 1.5e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 ORKTRNGFRVPL 13
DB      326 OOKLRSVFRVPI 337

```

```

RESULT 15
US-10-133-056-2
; Sequence 2, Application US/10133056
; Publication No. US2003008339A1
; GENERAL INFORMATION:
; APPLICANT: Tamura, Masaki
; TITLE OF INVENTION: Compositions and Methods of Treating Colorectal Polyps and Cancer
; FILE REFERENCE: Vanderbilt Ref. No. US2003008339A1 V00112; Our Ref. No. US200300833
; CURRENT APPLICATION NUMBER: US/10/133,056
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 363
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-133-056-2

```

```

Query Match      45.1%; Score 37; DB 15; Length 363;
Best Local Similarity 58.3%; Pred. No. 1.5e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 ORKTRNGFRVPL 13
DB      326 OOKLRSVFRVPI 337

```

```

Search completed: July 30, 2003, 16:35:14
Job time : 12.2381 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 31.5238 Seconds

(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-7

Perfect score: 82
Sequence: 1 TQKTRNGFRVPLARB 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_16Jun03:*

1: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDSI/gcgcdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	16	21	AA188342
2	82	100.0	712	12	AA112124
3	43	52.4	525	22	ABG05707
4	42	51.2	715	23	ABG69521
5	42	51.2	1394	23	ABP66310
6	41	50.0	81	23	ABR02245
7	41	50.0	876	23	ABR91807
8	40	48.8	101	22	AAW06693
9	40	48.8	294	21	AA35223

10	40	48.8	339	21	AA35222	Zea mays protein f
11	40	48.8	391	21	AA35221	Zea mays protein f
12	40	48.8	1333	22	AA35294	C glutamicum prote
13	39	47.6	62	22	AAU14545	Human novel protei
14	39	47.6	95	22	ABG07345	Novel human diagno
15	39	47.6	243	22	ABG02544	Novel human diagno
16	39	47.6	513	22	ABG03056	Novel human diagno
17	39	47.6	245	21	AAV77952	A. thaliana enviro
18	39	47.6	659	22	ABG00999	Novel human diagno
19	39	47.6	687	22	ABG19134	Novel human diagno
20	39	47.6	1156	20	AAV23754	Human nucleotide p
21	39	47.6	1325	22	ABG49735	Human liver peptid
22	39	47.6	1325	22	ABB29725	Peptide #2376 enco
23	39	47.6	1325	22	ABB34897	Peptide #2403 enco
24	39	47.6	1325	22	ABB20314	Protein #2313 enco
25	39	47.6	1325	22	AAW55707	Human brain expres
26	39	47.6	1325	22	AAW68085	Human bone marrow
27	39	47.6	1325	22	AAW03645	Peptide #2327 enco
28	39	47.6	1325	22	ABG37612	Human peptide enco
29	39	47.6	1684	22	ABE12028	Human dJ1042K10.4
30	39	47.6	1684	22	AAW79579	Human protein SEQ
31	39	47.6	1723	22	AAW78595	Human protein SEQ
32	38	46.3	69	22	AAU58868	Propionibacterium
33	38	46.3	124	21	AAV87153	Human secreted pro
34	38	46.3	124	21	AAE06130	Human gene 58 enco
35	38	46.3	124	22	AAE06185	Human gene 58 enco
36	38	46.3	124	23	ABG33953	Human secreted pro
37	38	46.3	124	23	ABG34009	Human secreted pro
38	38	46.3	125	21	AAV87153	Human secreted pro
39	38	46.3	159	22	ABG16246	Novel human diagno
40	38	46.3	172	23	AAV04964	M. tuberculosis an
41	38	46.3	172	23	ABU05945	M. tuberculosis an
42	38	46.3	186	20	AAV04965	Mycobacterium spec
43	38	46.3	225	22	AAW39172	Human polypeptide
44	38	46.3	263	23	AAE14443	Human drug metabol
45	38	46.3	330	21	AAV91506	Human secreted pro

ALIGNMENTS

RESULT 1
ID AA188342
XX AA188342: standard; peptide; 16 AA.

AC AA188342:
XX
XX
DT 14-JUL-2000 (first entry)
DB Peptide derived from an internal fragment of sheep whey.
KM Peptide production; biological fluid; milk; whey; blood;
KW antibacterial peptide; lactoferrin; antiviral; antitumour activity.
XX
XX Ovis sp.
OS
XX
XX
PN WO200015655-A1.
XX
XX 23-MAR-2000.
PD
PF 15-SEP-1999; 99WO-EP07002.
XX
XX 15-SEP-1998; 98EP-0203107.
PR 08-JUN-1999; 99EP-0201815.
PA (NIZO-) NIZO FOOD RES.
PI Visser S, Recio I;
DR WPI; 2000-271377/23.
XX Novel process for producing peptides with e.g. antimicrobial activity
PT from biological fluids such as milk, whey or blood comprises contacting

PT fluid with chromatographic medium to adsorb peptide domain of interest
 XX
 PS Claim 14; Page 22; 41pp; English.

CC This sequence represents a peptide derived from an internal fragment of
 CC sheep whey. The peptide is an example of a peptide with antibacterial
 CC activity that can be produced by the process of the invention. The
 CC invention relates to a process for producing peptides from biological
 CC fluids. The process comprises chromatography of the biological fluid, in
 CC situ hydrolysis of selectively bound peptides, washing to remove unbound
 CC peptide, and elution of the peptides of interest. The process is used for
 CC producing peptides from biological fluids, such as milk, whey or blood.
 CC For example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumour activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

SO Sequence 16 AA;

Query Match 100.0%; Score 82; DB 21; Length 16;
 Best Local Similarity 100.0%; Pred. No. 7.2e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TORKTRNGRVPRLAR 16
 DB 1 TORKTRNGRVPRLAR 16

RESULT 2
 AAR12124
 ID AAR12124 standard; Protein; 712 AA.

AC AAR12124;
 DT 02-AUG-1991 (first entry)

DE Bovine lactoperoxidase prepro enzyme.

KM LPO; autoimmune disease; antimicrobial; crosslinking agent;
 XM Immunassay; preservative.

OS Bos taurus.

FX Key Location/Qualifiers
 FT Peptide 1..26

FT /label= signal sequence

FT /note= "or 1..28"

FT Protein 101..712
 FT /label= mature LPO
 FT /note= "or 129..712"

MO9106639-A.

16-MAY-1991.

31-OCT-1990; 90WO-US06317.

03-NOV-1989; 89US-0431634.

(IDEO-) IDEON CORP.

Seilhammer JU, Dull TU;

WPI; 1991-164189/22.

N-PSDB; AAQ11842.

Recombinant lacto-peroxidase and deoxyribonucleic acid encoding
 it - used to treat auto-immune diseases, preserve prod. e.g.

PT cosmetics, and to crosslink ligands to proteins.
 XX
 PS Claim 1; Fig 1; 31pp; English.

CC The sequence was deduced from a cDNA library prepd. from a bovine
 CC milk gland. The sequence comprises 712 amino acids including the
 CC signal sequence and prosequence. Two putative signal peptidase
 CC cleavage sites and two possible alternative termini of the cleaved
 CC propeptide are indicated. The LPO protein is an antimicrobial
 CC useful for the treatment of infectious diseases and for preserv-
 CC tion of perishable goods such as foods, beverages, and cosmetics.
 CC It is active against fungi, viruses, protozoa and bacteria, and is
 CC also effective against malignancies, autoimmune disease and trans-
 CC plant rejection. It can also be used as an agent for crosslinking
 CC protein to ligands. See also AAR12562.

SO Sequence 712 AA;

Query Match 100.0%; Score 82; DB 12; Length 712;
 Best Local Similarity 100.0%; Pred. No. 3.8e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TORKTRNGRVPRLAR 16
 DB 179 TORKTRNGRVPRLAR 194

RESULT 3
 ABG05707
 ID ABG05707 standard; Protein; 525 AA.

AC ABG05707;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #5698.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

N-PSDB; AAS69894.

New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

Claim 20; SEQ ID No 36066; 103pp; English.

The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probe,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or

quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (ii) and its binding partners are useful in medical imaging of sites expressing (ii). (i) and (ii) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 525 AA;

Query Match 52.4%; Score 43; DB 22; Length 525;
Best Local Similarity 75.0%; Pred. No. 31;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
1 TORKTRNGFRVP 12
456 TORSTRLGVRVP 467

RESULT 4

ABP69521
ID ABP69521 standard; Protein; 715 AA.

ABP69521;

20-JAN-2003 (first entry)

Human polypeptide SEQ ID NO 1568.

Human; genome mapping; gene therapy; food supplement; virus; fungus; cell-proliferative disorder; neurodegenerative disease; bacterial; Parkinson's disease; Alzheimer's disease; autoimmune disease; multiple sclerosis; diabetes; genetic disorder; wound; burn; infection; arthritis; cystostatic; immunomodulator; nootropic; neuroprotective; antiparkinsonian; antidiabetic; immunosuppressive; dermatological; haemostatic; antiviral; fungicide; antibacterial; virucide; protozoacide; antiparasitic.

Homo sapiens.

WO200270539-A2.

12-SEP-2002.

05-MAR-2002; 2002WO-US05095.

05-MAR-2001; 2001US-0799451.

(HYSE-) HYSEQ INC.

Tang YT, Zhou P, Goodrich RM, Agundi V, Zhang J, Zhao QA, Ren F, Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M, Wehrman T, Wang J, Wang D, Drmanac RT;
WPI; 2002-759812/82.
N-PSDB; AB211738.

New polynucleotides comprising sequences assembled from expressed sequence tags (ESTs), useful for treating cell-proliferative, neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet or coagulation disorders -

Claim 9; SEQ ID NO 1568; 1012bp + Sequence listing; English.

The invention relates to an isolated polynucleotide (I) comprising a nucleotide sequence selected from any of 948 sequences (AB21119-AB212066) or their mature protein coding portion, active domain

coding protein or complementary sequences. The polynucleotides are useful for identifying expressed genes or for physical mapping of human genome. The encoded polypeptides (ABP6902-ABP69849) are useful as molecular weight markers, as a food supplement, for generating antibodies, in medical imaging, screening and diagnostic assays and for treating cell-proliferative disorders (cancer), neurodegenerative diseases (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid disorders, platelet or coagulation disorders, wound, burns, incision, ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal, parasitic), arthritis, etc.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 715 AA;

Query Match 51.2%; Score 42; DB 23; Length 715;
Best Local Similarity 53.3%; Pred. No. 65;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
1 TORKTRNGFRVPLAR 15
186 TFSRRRNGFLPLVR 200

RESULT 5

ABP66310
ID ABP66310 standard; Protein; 1394 AA.

ABP66310;

19-NOV-2002 (first entry)

Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:1054.

Bifidobacterium longum NCC2705; Bifidobacterium; bacterial; antidiarrhetic; antibacterial; inhibitor of Salmonella; detection; identification; lactic acid bacterium; diarrhoea; pathogenic bacteria; rotavirus; food composition; pharmaceutical composition.

Bifidobacterium longum.

EP1227152-A1.

31-JUL-2002.

30-JAN-2001; 2001EP-0102050.

30-JAN-2001; 2001EP-0102050.

(NEST) SOC PROD NESTLE SA.

WPI; 2002-668397/72.

Novel polynucleotide comprising Bifidobacterium genome sequence useful as a probe or primer for detecting and/or identifying Bifidobacterium longum in a biological sample -

Claim 3; SEQ ID 1054; 80bp; English.

The present invention describes a polynucleotide (I) comprising a sequence of a Bifidobacterium genome selected from the nucleotide sequences given in ABQ81842 and ABQ81843; or a sequence exhibiting at least 90% identity or which hybridises with the sequences given in ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a fusion protein, comprising a sequence selected from 1097 sequences given in ABP6258 to ABP6354 ligated in frame to a polynucleotide encoding a heterologous polypeptide. (II) has antidiarrhetic and antibacterial activities, and can be used as an inhibitor of Salmonella. (I) (which is a probe) is useful for the detection and/or identification of Bifidobacterium longum in a biological sample. A carrier containing the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618)

CC can be used for preventing and/or treating diarrhoea brought about by
 CC pathogenic bacteria and/or rotaviruses. The carrier is a food composition
 CC selected from milk, yogurt, curd, cheese, fermented milks, milk based
 CC fermented products, ice-creams, fermented cereal based products, milk
 CC based powders, infant formula, pet food or a pharmaceutical composition
 CC selected from tablets, liquid bacterial suspensions, dried oral
 CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.
 CC (1) is useful in DNA arrays or chips to carry out analysis of the
 CC expression of the Bifidobacterium gene. AB081844 to AB081850 represent
 CC Bifidobacterium related nucleotide sequences given in the Sequence
 CC listing from the present invention but not mentioned further within the
 CC specification.
 CC N.B. The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied by the
 CC European Patent Office.

CC XX Sequence 1394 AA;
 SQ

Query Match 51.2%; Score 42; DB 23; Length 1394;
 Best Local Similarity 61.5%; Pred. No. 1.3e+02;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 KTRNGFRVPLARE 16
 : ||||| :
 Db 694 QTVNGFRTPARX 706

RESULT 6
 ID ABP02245
 XX ABP02245 standard; Protein; 81 AA.
 XX AC
 XX ABP02245;
 DT 25-JUN-2002 (first entry)

DE Human ORFX protein sequence SEQ ID NO:4472.
 XX
 XX Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
 XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 XX hypertension; hypothyroidism; cholesterol ester storage disease;
 XX immune deficiency; immune disorder; infectious disease;
 XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 XX myasthenia gravis.
 XX
 OS Homo sapiens.
 XX
 XX WO200192523-A2.
 XX
 XX PD 06-DEC-2001.
 XX
 XX PF 29-MAY-2001; 2001WO-US10836.
 XX
 XX PR 30-MAY-2000; 2000US-206132P.
 XX PR 29-AUG-2000; 2000US-228716P.
 XX
 XX PA (CURA-) CURAGEN CORP.
 XX
 XX PI Shimkets RA, Leach MD;
 XX
 XX DR WPI; 2002-106308/14.
 XX DR N-PSDB; ABN17997.
 XX
 XX PT Novel human polypeptides and polynucleotides useful for diagnosing,
 XX PT preventing and treating cardiovascular disease, neurodegenerative,
 XX PT hyperproliferative disorders and autoimmune disorders -
 XX
 XX PS Disclosure; SEQ ID 4472; 1037bp; English.
 CC
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX

CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC XX Sequence 81 AA;
 SQ

Query Match 50.0%; Score 41; DB 23; Length 81;
 Best Local Similarity 80.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 NGFNVPPLARE 16
 ||||| :
 Db 38 NGFNVPPLASE 47

RESULT 7
 ID ABB91807
 XX ABB91807 standard; Protein; 876 AA.
 XX AC
 XX ABB91807;
 DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 1018.
 XX
 XX DE Herbicidally active polypeptide SEQ ID NO 1018.
 XX
 XX KM Herbicidal; plant; agriculture; herbicide.
 XX
 OS Arabidopsis thaliana.
 XX
 XX OS
 XX WO200210210-A2.
 XX
 XX PD 07-FEB-2002.
 XX
 XX PF 28-AUG-2001; 2001WO-EP09892.
 XX
 XX PR 28-AUG-2001; 2001WO-EP09892.
 XX
 XX PA (FARB) BAYER AG.
 XX
 XX PI Tietjen K, Weidler M;
 XX
 XX DR WPI; 2002-269010/31.
 XX
 XX PT Identifying plant target proteins for herbicidally active compounds,
 XX PT comprising aligning and comparing nucleic acid or amino acid sequences
 XX PT from plant with nucleic acid or amino acid sequences from non-plant
 XX PT organisms -
 XX
 XX PS Claim 5; SEQ ID NO 1018; 261bp + Sequence listing; English.
 CC
 CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140891.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141882.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142970.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147382.
PR 05-AUG-1999; 99US-0147380.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147483.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154019.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161923.
PR 28-OCT-1999; 99US-0161992.
PR 29-OCT-1999; 99US-0162142.

Query Match 48.8%; Score 40; DB 21; Length 294;
Best Local Similarity 56.2%; Pred. No. 59;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 TQKTRNGRFLPARE 16
Db 5 TTRKTRDPPATIKARE 20

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 48.8%; Score 40; DB 21; Length 339;
Best Local Similarity 56.2%; Pred. No. 68;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 1 TORTRNGFVPLARE 16
Db 50 TTRKTRDPFAIRKARE 65

RESULT 11
AAG35221
ID AAG35221 standard; Protein; 391 AA.
XX AC AAG35221;
XX,

DT 18-OCT-2000 (first entry)
XX Zea mays protein fragment SEQ ID NO: 42995.
DE Protein identification; signal transduction pathway; metabolic pathway;
XX hybridization assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142290.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0143277.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 48.8%; Score 40; DB 21; Length 391;
Best Local Similarity 56.2%; Pred. No. 79;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TORXTRNGFRVPLARE 16
Db 102 TTRKTRDPPAIRKARE 117

RESULT 12

AAG90294

ID AAG90294 strand: Protein; 1333 AA.

XX AAG90294;

XX 26-SEP-2001 (first entry)

XX C glutamicum protein fragment SEQ ID NO: 4048.

DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis.

XX

OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
DR WPI, 2001-376931/40.
DR N-PSDB; AAB65513.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
PS Claim 17; SEQ ID NO: 4048; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 1333 AA;
XX
Query Match 48.8%; Score 40; DB 22; Length 1333;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 5 TRNGFRVPLARE 16
Db 903 TRQGVRFVPAAE 914
XX
RESULT 13
AAU14545
ID AAU14545 standard; Protein; 62 AA.
XX
AC AAU14545;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human novel protein #416.
XX
KW Human: novel protein; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;
KW anticonvulsant; antiallergic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
PN WO200155437-A2.

XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02623.
XX
PR 25-JAN-2000; 2000US-0491404.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
PI WPI, 2001-451939/48.
DR N-PSDB; AAS22850.
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
PS Example 4; Page 866; 894pp; English.
XX
CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/ elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
XX
SQ Sequence 62 AA;
XX
Query Match 47.6%; Score 39; DB 22; Length 62;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 4 KTRNGFRVP 12
Db 32 ETRDGFRIIP 40
XX
RESULT 14
ABG07345
ID ABG07345 standard; Protein; 95 AA.
XX
AC ABG07345;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #7336.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.

```

PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT.
XX
XX MPI; 2001-639362/73.
XX
XX N-PSDB; AAS71352.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 20; SEQ ID No 37704; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 95 AA;
SQ
XX
XX Query Match 47.6%; Score 39; DB 22; Length 95;
XX Best Local Similarity 77.8%; Pred. No. 27;
XX Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 4 KTRNGFRVP 12
XX |||||
XX | | | |
XX | | | |
XX Db 77 KTRNGFRIP 85
XX
XX RESULT 15
XX ABG02544
XX ID ABG02544 standard; Protein; 243 AA.
XX
XX AC ABG02544;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #2535.
XX
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX

```

```

PF 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX MPI; 2001-639362/73.
XX
XX N-PSDB; AAS66731.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 20; SEQ ID No 32903; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 243 AA;
SQ
XX
XX Query Match 47.6%; Score 39; DB 22; Length 243;
XX Best Local Similarity 77.8%; Pred. No. 73;
XX Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 4 KTRNGFRVP 12
XX |||||
XX | | | |
XX | | | |
XX Db 225 KTRNGFRIP 233
XX
XX Search completed: July 30, 2003, 16:23:30
XX Job time : 33.5238 secs
XX

```

THIS PAGE BLANK (U&PTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:36 ; Search time 13.5714 Seconds
(without alignments)
152.115 Million cell updates/sec

Title: US-09-787-070-8

Perfect score: 47
Sequence: 1 APRKNVRW 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	33	6 Q9TR80	Q9tr80 ovls aries
2	41	87.2	266	5 Q9BRW0	Q9brw0 caenorhabd
3	38	80.9	48	6 Q77558	Q77558 sus scrofa
4	38	80.9	704	6 Q8MMN8	Q8mmn8 sus scrofa
5	38	80.9	1072	5 Q8T104	Q8t104 bombyx mori
6	37	78.7	314	11 Q8K2X0	Q8k2x0 mus musculu
7	37	78.7	555	4 Q96N65	Q96n65 homo sapien
8	37	78.7	555	11 Q8BMA3	Q8bma3 mus musculu
9	37	78.7	1126	5 Q9BPR0	Q9bpr0 bombyx mori
10	37	78.7	1712	5 Q8ICAI	Q8icai plasmodium
11	36	76.6	166	5 Q8SRC9	Q8src9 encephalito
12	36	76.6	238	5 Q8SYH1	Q8syh1 drosophila
13	36	76.6	1051	12 Q66108	Q66108 citrus leaf
14	35	74.5	345	5 P91109	P91109 caenorhabd
15	35	74.5	372	2 Q05723	Q05723 bacteroides
16	35	74.5	379	16 Q98CW3	Q98cw3 rhizobium 1

17	35	74.5	379	16 Q8YQ02	Q8yq02 bruceella me
18	35	74.5	379	16 Q8G161	Q8g161 bruceella su
19	35	74.5	404	16 Q9W82	Q9w82 straphylococ
20	35	74.5	404	16 Q8CO90	Q8co90 straphylococ
21	35	74.5	484	17 Q8TNS2	Q8tns2 methanosaer
22	35	74.5	496	15 Q99412	Q99412 human immun
23	35	74.5	2513	12 Q87644	Q87644 sindbis vir
24	34	72.3	96	5 Q81J88	Q81j88 plasmodium
25	34	72.3	118	10 Q9AS17	Q9as17 cryza sativ
26	34	72.3	122	12 Q64873	Q64873 human adeno
27	34	72.3	256	4 Q96MX8	Q96mx8 homo sapien
28	34	72.3	351	12 Q8B8T6	Q8b8t6 human adeno
29	34	72.3	356	12 Q8UY81	Q8uy81 simian aden
30	34	72.3	380	16 Q9F2R4	Q9f2r4 streptomyc
31	34	72.3	410	17 Q8TZP3	Q8tzp3 pyrococcus
32	34	72.3	445	10 Q8H086	Q8h086 cryza sativ
33	34	72.3	462	15 Q72608	Q72608 human immun
34	34	72.3	491	16 Q8FN57	Q8fn57 corynebacte
35	34	72.3	500	15 Q9WQ10	Q9wq10 human immun
36	34	72.3	518	16 Q8KB21	Q8kb21 chlorobium
37	34	72.3	1032	13 Q8AX24	Q8ax24 brachydanio
38	34	72.3	1277	12 Q9JG17	Q9jg17 sorghum chl
39	33	70.2	85	12 Q98433	Q98433 paramacium
40	33	70.2	164	5 Q17570	Q17570 caenorhabd
41	33	70.2	183	16 Q9KYJ3	Q9kyj3 streptomyc
42	33	70.2	185	5 Q94783	Q94783 trypanosoma
43	33	70.2	202	16 Q934V8	Q934v8 salmonella
44	33	70.2	227	5 Q18487	Q18487 caenorhabd
45	33	70.2	244	12 Q41868	Q41868 asparagus v

ALIGNMENTS

RESULT 1					
Q9TR80		PRELIMINARY;	PRT;	33 AA.	
AC Q9TR80;					
DT 01-MAY-2000 (TREMBLrel. 13, Created)					
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)					
DE Lactoferrin (Fragment).					
OS Ovis aries (Sheep).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;					
OC Bovidae; Caprinae; Ovis.					
OX NCBI_TaxID=9940;					
RP [1]					
RP SEQUENCE.					
RX MEDLINE=95127729; PubMed=7827104;					
RL Qian Z.Y., Jolles P., Migliore-Samour D., Fiat A.M.;					
RL Biochim. Biophys. Acta 1243:25-32(1995).					
DR HSP; O77698; ICE2.					
DR InterPro; IPR001156; Transferrin.					
DR Pfam; PF00405; transferrin; 1.					
SO SEQUENCE 33 AA; 3914 MW; D1904CAF15A73961 CRC64;					
Query Match	100.0%;	Score 47;	DB 6;	Length 33;	
Best Local Similarity	100.0%;	Pred. No. 0.016;			
Matches 8; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY 1 APRKNVRW 8					
DB 1 APRKNVRW 8					
RESULT 2					
Q9BKW0		PRELIMINARY;	PRT;	266 AA.	
ID Q9BKW0;					
DT 01-JUN-2001 (TREMBLrel. 17, Created)					
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)					
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)					

DE Hypothetical protein Y92H12BL.4.
 GN Y92H12BL.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Poloderrinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA Waterston R.;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Tin-Mollam A., Woldmann P., Courtney L.;
 RT "The sequence of C. elegans cosmid Y92H12BL.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC084268; AAK19005.3; -
 KW Hypothetical protein.
 SQ SEQUENCE 266 AA; 30750 MW; 4CA232BD999AAB5 CRC64;

Query Match 87.2%; Score 41; DB 5; Length 266;
 Best Local Similarity 75.0%; Pred. No. 2.3;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 APRKXVW 8
 Db 109 APRKXVW 116

RESULT 3
 ID 077558 PRELIMINARY; PRT; 48 AA.
 AC 077558;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Lactoferrin (Fragment).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxId=9823;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98332734; PubMed=9666128;
 RA Wang S.-R., Lin J., Cheng I.C., Lin T.Y.;
 RT "Characterization and functional analysis of the porcine lactoferrin
 RT gene promoter.";
 RL Gene 215:203-212(1998).
 DR EMBL; AF044256; AAC34369.1; -
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; Transferrin; 1.
 FT NON TER 48
 PT 48
 SQ SEQUENCE 48 AA; 5569 MW; 28403BE7DE144D78 CRC64;

Query Match 80.9%; Score 38; DB 6; Length 48;
 Best Local Similarity 75.0%; Pred. No. 1.7;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 APRKXVW 8
 Db 20 APRKXVW 27

RESULT 4

Q8WMN8
 ID Q8WMN8 PRELIMINARY; PRT; 704 AA.
 AC Q8WMN8;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Lactoferrin (Fragment).
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxId=9823;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Wang S.-R., Lin T.-Y., Weng C.-N.;
 RT Isolation and expression of porcine milk lactoferrin.";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 CC OF AN ANION, USUALLY BICARBONATE (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 DR EMBL; L77887; AAI40161.1; -
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; Transferrin; 2.
 DR PRINTS; PR00422; TRANSFERRIN.
 DR SMART; SMO0094; TR FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.
 KW Glycoprotein; Iron transport; Metal-binding; Transport.
 FT NON TER 1
 PT 1
 SQ SEQUENCE 704 AA; 77681 MW; 64EB769F7503CC32 CRC64;

Query Match 80.9%; Score 38; DB 6; Length 704;
 Best Local Similarity 75.0%; Pred. No. 26;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 APRKXVW 8
 Db 20 APRKXVW 27

RESULT 5
 ID 08T104 PRELIMINARY; PRT; 1072 AA.
 AC 08T104;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Projection-like protein.
 GN BMPROJECTIN.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
 OC Bombycidae; Bombyx.
 OX NCBI_TaxId=7091;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=50; TISSUE=Posterior silk gland;
 RA Koike Y., Shimada T., Suzuki M.G., Mita K., Abe H., Maeda S.,
 RA Oseegawa K., Delong P.J.;
 RT "Genomic sequence of 320kb containing a kettin orthologue on the Z
 RT chromosome in Bombyx mori.";
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=50; TISSUE=Posterior silk gland;
 RA Koike Y., Shimada T., Suzuki M.G., Mita K., Abe H., Maeda S.,
 RA Oseegawa K., Delong P.J.;
 RT "Genomic sequence of 320kb containing a kettin orthologue on the Z
 RT chromosome in Bombyx mori.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB079864; BAB5195.1; -
 DR EMBL; AB090307; BAC10616.1; -
 DR InterPro; IPR003961; FN_111.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003006; IG_MHC.
 DR Pfam; PF00041; fn3; 2.
 DR Pfam; PF00047; Ig; 1.
 DR PROSITE; P550835; IG_LIKE; 2.
 SQ SEQUENCE 1072 AA; 122004 MW; C7052490D28BE29E CRC64;

Query Match
 Best Local Similarity 80.9%; Score 38; DB 5; Length 1072;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRKNVRW 8
 DB 14 PRKNVRW 20

RESULT 6
 Q8K2KO PRELIMINARY; PRT; 314 AA.

AC Q8K2KO; 01-OCT-2002 (TREMblrel. 22, Created)
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Similar to maguin-2, maguin-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strusberg R.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC031194; AAH31194.1; -
 DR InterPro; IPR001478; PDZ; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR PROSITE; P550106; PDZ; 1.
 SQ SEQUENCE 314 AA; 34995 MW; 7A263D3AF0E51F79 CRC64;

Query Match
 Best Local Similarity 78.7%; Score 37; DB 11; Length 314;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 DB 61 APLKNLRW 68

RESULT 7
 Q96N65 PRELIMINARY; PRT; 555 AA.

AC Q96N65; 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Hypothetical protein FLJ13149.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homindaes; Homo.
 OX NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ishibashi T., Kaneshi K., Yosida M., Watanabe S., Ishida S., Ono Y.,
 RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Masuno K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Magatsuna M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.;
 RN "NEDO human cDNA sequencing project.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

DR EMBL; AK055911; BAB71044.1; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001601; SAM.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00536; SAM; 1.
 DR SMART; SM00228; PDZ; 1.
 DR SMART; SM00454; SAM; 1.
 DR PROSITE; P550106; PDZ; 1.
 DR PROSITE; P550105; SAM_DOMAIN; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 555 AA; 61891 MW; 6EF900E174D91C08 CRC64;

Query Match
 Best Local Similarity 78.7%; Score 37; DB 4; Length 555;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 DB 302 APLKNLRW 309

RESULT 8
 Q8BMA3 PRELIMINARY; PRT; 555 AA.

AC Q8BMA3; 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Weakly similar to membrane-associated guanylate kinase-interacting
 DE protein 2 MAGUIN-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6J; TISSUE=Mullerian duct includes surrounding region;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573 (2002).
 DR EMBL; AK033015; BAC28127.1; -
 SQ SEQUENCE 555 AA; 61869 MW; 341AF9A4425904F2 CRC64;

Query Match
 Best Local Similarity 78.7%; Score 37; DB 11; Length 555;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 DB 302 APLKNLRW 309

RESULT 9
 Q9BPRO PRELIMINARY; PRT; 1126 AA.

AC Q9BPRO; 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Receptor type guanylyl cyclase (EC 4.6.1.2) (Guanylate
 DE cyclase).
 GN BM-GCY.
 OS Bombyx mori (Silk moth).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditytaria; Bombycoidea;
 OC Bombycidae; Bombyx.
 OX NCB1_TaxID=7091;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Shintarou T., Shigeki S., Taeko S., Yasuhisa E., Takaaki N.,
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21376156; PubMed=11483433;
 RA Tanoue S., Sumida S., Suetsugu T., Endo Y., Nishiohka T.;
 RT "Identification of a receptor type guanylyl cyclase in the antennal
 lobe and antennal sensory neurons of the silkworm, Bombyx mori.";
 RL Insect Biochem Mol Biol 31:971-979(2001)
 CC -1- CATALYTIC ACTIVITY: GTP = 3',5'-CYCLIC GMP + DIPHOSPHATE.
 CC -1- CATALYTIC ACTIVITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL
 CC CYCLASE FAMILY.
 DR EMBL; AB047558; BAB32672.1; -.
 DR HSSP; P26769; 1AB8.
 DR InterPro; IPR001828; ANF_receptor.
 DR InterPro; IPR001054; G_cyclase.
 DR InterPro; IPR001993; Mitoch_carrier.
 DR InterPro; IPR001170; Ntpep_receptor.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF01094; ANF_receptor; 1.
 DR Pfam; PF00211; guanylate_cyc; 2.
 DR Pfam; PF00069; pk_nase; 1.
 DR PRINTS; PR00255; NATPEPTIDER.
 DR PRINTS; PR00109; TYRKINASE.
 DR PRODOM; PD000001; Prot_kinase; 1.
 DR SMART; SM00044; CYCC; 1.
 DR PROSITE; PS00452; GUANYLYLATE_CYCLASES_1; 1.
 DR PROSITE; PS50125; GUANYLYLATE_CYCLASES_2; 1.
 DR PROSITE; PS500215; MITOCH_CARRIER; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR ATP-binding; Kinase; Lyase; Transferase; Tyrosine-protein kinase;
 KW CGMP synthesis;
 SQ SEQUENCE 1126 AA; 126212 MW; F6989C5B0E0A9488 CRC64;

Query Match
 Best Local Similarity 78.7%; Score 37; DB 5; Length 1126;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 APRKXVW 8
 DB 1077 APKXLRW 1084

RESULT 10
 Q8ICAI PRELIMINARY; PRT; 1712 AA.
 ID Q8ICAI;
 AC Q8ICAI;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN MAL6P1.157.
 OS Plasmodium falciparum (Isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,
 RA Bertman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL844505; CAD50541.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 1712 AA; 198276 MW; B2DB83661670036C CRC64;

Query Match
 Best Local Similarity 78.7%; Score 37; DB 5; Length 1712;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 PRKXVW 8
 DB 1111 11

DB 256 PRKXVW 262
 RESULT 11
 ID Q8SRC9 PRELIMINARY; PRT; 166 AA.
 AC Q8SRC9;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
 DE Protein involved in RETRIEVAL of ER membrane proteins from the early
 DE golgi compartment.
 GN EC008.0700.
 OS Eucephalitozoan cuniculi.
 OC Eukaryota; Fungi; Microsporidia; Unikaryontidae; Eucephalitozoon.
 NCBI_TaxID=6035;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RA Genoscope;
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GB-M1;
 RX MEDLINE=21576510; PubMed=11719806;
 RA Katinka M.D., Duprat S., Cornillon E., Metenier G., Thomarat F.,
 RA Prentier G., Barbe V., Peyretallade E., Brottier P., Mincker P.,
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
 RA Weissendbach J., Vivares C.P.;
 RT "Genome sequence and gene compaction of the eukaryote parasite
 RT Eucephalitozoon cuniculi.";
 RL Nature 414:450-453(2001).
 DR EMBL; AL590448; CAD26375.1; -.
 DR InterPro; IPR004932; Ret1.
 DR Pfam; PF03248; Ret1; 1.
 SQ SEQUENCE 166 AA; 19690 MW; 51D58C48E03F3A31 CRC64;

Query Match
 Best Local Similarity 76.6%; Score 36; DB 5; Length 166;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 APRKXVW 8
 DB 15 APRKXVW 22

RESULT 12
 Q8SYH1 PRELIMINARY; PRT; 238 AA.
 ID Q8SYH1;
 AC Q8SYH1;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE RE62711P (CG32442-PA).
 GN BCDA:RB62711 OR CG32442.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
 RA Champagne M., Chavez C., Dorsett V., Dresner D., Fartin D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nuno J., Pacleb J., Paragae V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celisner S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Chapple M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Gabor G.L.,
 RA Abiri J.F., Agopyani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benson P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher S., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Idegawa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Modyarty C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murzyn D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
 RA Reibert K., Remington K., Saunders R.D., Schebler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith I.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasserman D.A., Weissstock G.M., Weissbach J.,
 RA Williams S.M., Woodgett, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 [3]
 RP SEQUENCE FROM N.A.
 RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amaratides P.G., Brandon R.C., Rogers Y.,
 RA Banson J., An H., Baldwin D., Banson J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Chapple M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.B., Doyle C., Drenek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A., J.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Idegawa C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Schebler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celinker S.E.,
 RA Clamp M., Drysdale R., Emert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [6]
 RP SEQUENCE FROM N.A.
 RA Flybase;

RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY071557; AAL49179.1; -
 DR EMBL; AE003594; AAL12176.1; -
 DR Flybase; FBgn0047079; BCDNA:R862711.
 DR InterPro; IPR000345; Cytochrome_bcd.
 DR Pfam; PF04161; Arv1, 1.
 DR PROSITE; PS00190; CYTOCHROME_C_1.
 SQ SEQUENCE 238 AA; 27748 MW; 4A57CE830C4B7641 CRC64;
 Query Match 76.6%; Score 36; DB 5; Length 238;
 Best Local Similarity 75.0%; Pred.No. 22;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 APRKAVRW 8
 DB 204 ATKRLRW 211
 RESULT 13
 ID 066108 PRELIMINARY; PRT; 1051 AA.
 AC 066108;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Citrus leaf rugose ilavirius putative viral replicase.
 OC Citrus leaf rugose virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae;
 OC Ilavirius.
 OX NCBI_TaxID=37126;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96112216; PubMed=8847536;
 RT Scott S.W., Ge X.,
 RT "The nucleotide sequence of citrus leaf rugose virus RNA 1.";
 RT J. Gen. Virol. 76:323-328(1995).
 DR EMBL; U23715; AAC49126.1; -
 DR InterPro; IPR000606; Viral_helicase1.
 DR InterPro; IPR002588; V_methyltransferase.
 DR Pfam; PF01443; Viral_helicase1.1.
 DR Pfam; PF01600; Vmethyltransferase.1.
 SQ SEQUENCE 1051 AA; 118326 MW; A686A74981D9859F CRC64;
 Query Match 76.6%; Score 36; DB 12; Length 1051;
 Best Local Similarity 85.7%; Pred.No. 1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 PRKAVRW 8
 DB 231 PLRNVRW 237
 RESULT 14
 ID P91109 PRELIMINARY; PRT; 345 AA.
 AC P91109;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypochemical 40.6 kDa protein.
 GN C32B5.16.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Felodertinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode *C. elegans*: a platform for
 RT investigating biology. The *C. elegans* Sequencing Consortium.";
 RL Science 282:2012-2018(1998).

Search completed: July 30, 2003, 16:29:30
 Job time : 16.5714 secs

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Sheet P.;
 RT "The sequence of C. elegans coamid C32B5";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Waterston R.;
 RT "Direct Submission";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U80843; AAB37971.1; -
 DR Wormpep; C32B5.16; CE08524.
 KW Hypothetical protein.
 SQ SEQUENCE 345 AA; 40557 MW; 1BC8264AE08F32B4 CRC64;

Query Match 74.5%; Score 35; DB 5; Length 345;
 Best Local Similarity 83.3%; Pred. No. 53;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RKXVW 8
 Db 100 RKXVW 105

RESULT 15
 ID 005723 PRELIMINARY; PRT; 372 AA.
 AC 005723;
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Integrase.
 GN INT.
 OS Bacteroides fragilis.
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Bacteroidaceae; Bacteroides.
 OX NCBI_Taxid=817;
 RN [1]
 RP SEQUENCE OF 1-224 FROM N.A.
 RC TRANSPOSON-Tn4555;
 RX MEDLINE=97252504; PubMed=9098073;
 RA Tribble G.D., Parker A.C., Smith C.J.;
 RT "The Bacteroides mobilizable transposon Tn4555 integrates by a site-specific recombination mechanism similar to that of the gram-positive bacterial element Tn916";
 RT J. Bacteriol. 179:2731-2739(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-Tn4555;
 RX MEDLINE=99343846; PubMed=10413660;
 RA Tribble G.D., Parker A.C., Smith C.J.;
 RT "Genetic structure and transcriptional analysis of a mobilizable, antibiotic resistance transposon from Bacteroides";
 RT Plasmid 42:1-12(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TRANSPOSON-Tn4555;
 RA Tribble G.D., Parker A.C., Smith C.J.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U75371; AAB53787.2; -
 DR InterPro; IPR002104; Phage integrase.
 DR Pfam; PF00589; Phage integrase; 1.
 SQ SEQUENCE 372 AA; 43472 MW; F2F08A8F31ACD04D CRC64;

Query Match 74.5%; Score 35; DB 2; Length 372;
 Best Local Similarity 62.5%; Pred. No. 57;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 APRXVW 8
 Db 104 ADRXVW 111

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:10:10 ; Search time 2.61905 Seconds

(without alignments)
143.645 Million cell updates/sec

Title: US-09-787-070-8

Perfect score: 47
Sequence: 1 APRKXVRW 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	100.0	708	TRFL_BOVIN	P24627 boe taurus
2	47	100.0	708	TRFL_BOBBU	O77698 bubalus bub
3	47	100.0	708	TRFL_CAPHI	Q29477 capra hircu
4	42	89.4	695	TRFL_HORSE	O77811 equus cabal
5	38	80.9	704	TRFL_PIG	P14632 sus scrofa
6	35	74.5	2512	POLN_STNDV	P03317 sinchis vir
7	34	72.3	347	VCOM_ADE12	P36717 human adeno
8	34	72.3	358	VCOM_ADE40	P48753 human adeno
9	34	72.3	368	VCOM_ADE05	P24938 human adeno
10	34	72.3	369	VCOM_ADE02	P03267 human adeno
11	34	72.3	439	RT30_HUMAN	O9N929 homo sapien
12	33	70.2	166	LITR_HUMAN	P05451 homo sapien
13	33	70.2	166	LITR_HUMAN	P48304 homo sapien
14	33	70.2	194	RISB_APCFU	O27965 archaeoglob
15	33	70.2	326	BTUC_SALTY	O82615 salmoneila
16	33	70.2	326	BTUC_SALTY	O82615 salmoneila
17	33	70.2	382	YHME_YEAST	P32793 saccharomyc
18	33	70.2	404	OAL_HUMAN	P51810 homo sapien
19	33	70.2	694	LCF3_YEAST	P39002 saccharomyc
20	33	70.2	1451	EM30_AAVTH	Q42310 arabidopsis
21	33	70.2	3726	TRX_DROME	P20659 drosophila
22	32	68.1	96	YS98_MYCLE	Q33024 mycobacteri
23	32	68.1	442	RT30_MOUSE	O9D690 mus musculu
24	32	68.1	538	CIMA_PAT	O9J184 ratuus norv
25	32	68.1	694	LCF4_YEAST	P47912 saccharomyc
26	32	68.1	741	CH12_YEAST	P49556 saccharomyc
27	32	68.1	2236	PYRI_DROME	P05990 drosophila
28	31.5	67.0	510	YOT3_CAEEL	P34649 caenorhabdi
29	31	66.0	118	VE4_HPV13	Q02655 streptomyce
30	31	66.0	124	Y212_STRCO	O54625 bacillus su
31	31	66.0	189	MAP_BACSU	O02169 bacillus su
32	31	66.0	198	AICD_HUMAN	O99X27 mus musculu
33	31	66.0	198	AICD_MOUSE	O99X27 mus musculu

ALIGNMENTS

RESULT 1	ID	TRFL_BOVIN	STANDARD:	PRT:	708 AA.
AC	P24627	Q29629; Q9MZV3;			
DT	01-MAR-1992	(Rel. 21, Created)			
DT	01-OCT-1993	(Rel. 27, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Lactotransferrin precursor (Lactoferrin) [Contains: Lactoferricin B (LFCIN B)].				
GN	LTF.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_Taxid=9913;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Submaxillary gland;				
RX	MEDLINE=91160550; PubMed=2001696;				
RA	Pierce A., Colavizza D., Benalissa M., Maes F., Tartar A.,				
RA	Montreuil J., Spik G.;				
RT	"Molecular cloning and sequence analysis of bovine lactotransferrin.";				
RL	Eur. J. Biochem. 196;177-184(1991).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92028986; PubMed=1718281;				
RA	Goodman R.E., Schanbacher F.L.;				
RT	"Bovine lactoferrin mRNA: sequence, analysis, and expression in the				
RT	mammary gland.";				
RL	Biochem. Biophys. Res. Commun. 180:75-84(1991).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Lung;				
RT	Teaag T.C., Burns D.K., Wang F., Pan Y.C.E., Schmidt A.M., Stern D.;				
RT	"Cloning of a 80-kD advanced glycosylation end product (AGE) binding				
RL	protein from bovine lung.";				
RN	FASBB J. 6:233-233(1991).				
RP	[4]				
RC	SEQUENCE FROM N.A.				
RX	TISSUE=Blood, and Mammary gland;				
RA	MEDLINE=9426164; PubMed=8206385;				
RT	Seylert H.-M., Tuckoritz A., Interthal H., Koczan D., Hobom G.;				
RT	"Structure of the bovine lactoferrin-encoding gene and its promoter.";				
RN	Gene 143:265-269(1994).				
RP	[5]				
RC	SEQUENCE FROM N.A.				
RA	Nakamura T., Shimazaki K., Yagi Y., Watanabe A.;				
RL	Submitted (Jul-2000) to the EMBL/Genbank/DBJ databases.				
RN	[6]				
RP	SEQUENCE OF 20-59.				
RX	MEDLINE=90031466; PubMed=2805645;				
RA	Rejman J.J., Hegarty H.M., Hurley W.L.;				
RT	"Purification and characterization of bovine lactoferrin from				
RT	secretions of the involuting mammary gland: identification of				
RT	multiple molecular weight forms.";				
RL	Comp. Biochem. Physiol. 93B:929-934(1989).				

Q8x417 escherichia
P06609 escherichia
Q91427 homo sapien
P23065 bluetongue
P10350 bluetongue
Q9x499 vldrio chol
P39319 escherichia
Q8wx42 homo sapien
P47867 mus musculu
P47868 ratuus norv
Q9umh8 mesocistigma
Q8x9m0 buchnera ap

[7]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 RX MEDLINE=98062367; PubMed=9398529;
 RA Moore S.A., Anderson B.F., Groom C.R., Haridas M., Baker E.N.;
 RT "Three-dimensional structure of ferric bovine lactoferrin at 2.8-A
 resolution.";
 RL J. Mol. Biol. 274:222-236(1997).
 RN (8)
 RP STRUCTURE BY NMR OF 36-60.
 RX MEDLINE=98190007; PubMed=9521752;
 RA Hwang P.M., Zhou N., Shan X., Arrowsmith C.H., Vogel H.J.;
 RT "Three-dimensional solution structure of lactoferricin B, an
 antimicrobial peptide derived from bovine lactoferrin.";
 RL Biochemistry 37:4288-4298(1998).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE.
 CC -1- FUNCTION: LACTOFERRICIN B IS AN ANTIMICROBIAL PEPTIDE.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X57084; CAA40366.1; -;
 DR EMBL; M63502; AAA30617.1; -;
 DR EMBL; L08604; AAA30609.1; -;
 DR EMBL; L19983; AAA21722.1; -;
 DR EMBL; L19982; AAA21722.1; JOINED.
 DR EMBL; L19983; AAA21722.1; JOINED.
 DR EMBL; L19984; AAA21722.1; JOINED.
 DR EMBL; L19985; AAA21722.1; JOINED.
 DR EMBL; L19986; AAA21722.1; JOINED.
 DR EMBL; L19988; AAA21722.1; JOINED.
 DR EMBL; L19989; AAA21722.1; JOINED.
 DR EMBL; L19990; AAA21722.1; JOINED.
 DR EMBL; L19991; AAA21722.1; JOINED.
 DR EMBL; L19992; AAA21722.1; JOINED.
 DR EMBL; AB046664; BAB03470.1; -;
 DR PIR; I45919; TRBOU.
 DR PDB; 1BLF; 03-DEC-97.
 DR PDB; 1LFC; 18-NOV-98.
 DR GlycoSuiteDB; P24627; -;
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR PRINTS; PRO0422; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal; Antibiotic; 3D-structure.
 FT SIGNAL 1 19
 FT CHAIN 20 708 LACTOTRANSFERRIN.
 FT PEPTIDE 36 60 LACTOFERRICIN B.
 FT REPEAT 20 363 1.
 FT REPEAT 364 708 2.
 FT DISULFID 28 64 BY SIMILARITY.
 FT DISULFID 38 55
 FT DISULFID 134 217 BY SIMILARITY.
 FT DISULFID 176 192 BY SIMILARITY.
 FT DISULFID 189 200 BY SIMILARITY.
 FT DISULFID 250 264 BY SIMILARITY.
 FT DISULFID 367 399 BY SIMILARITY.
 FT DISULFID 377 390 BY SIMILARITY.
 FT DISULFID 424 703 BY SIMILARITY.

FT DISULFID 444 666 BY SIMILARITY.
 FT DISULFID 476 551 BY SIMILARITY.
 FT DISULFID 500 694 BY SIMILARITY.
 FT DISULFID 510 524 BY SIMILARITY.
 FT DISULFID 521 534 BY SIMILARITY.
 FT DISULFID 532 606 BY SIMILARITY.
 FT DISULFID 644 649 BY SIMILARITY.
 FT METAL 79 79 IRON 1.
 FT METAL 111 111 IRON 1.
 FT METAL 211 211 IRON 1.
 FT METAL 272 272 IRON 1.
 FT METAL 414 414 IRON 2.
 FT METAL 452 452 IRON 2.
 FT METAL 545 545 IRON 2.
 FT METAL 614 614 IRON 2.
 FT BINDING 140 140 ANION (POTENTIAL).
 FT BINDING 482 482 ANION (POTENTIAL).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 387 387 /FTid=CAR_000186.
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 564 564 /FTid=CAR_000197.
 FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .).
 FT CONFLICT 63 63 E -> A (IN REF. 4).
 FT CONFLICT 66 67 RA -> PG (IN REF. 2).
 FT CONFLICT 145 145 I -> V (IN REF. 1 AND 4).
 FT CONFLICT 164 165 LG -> PP (IN REF. 1).
 FT CONFLICT 254 264 C -> Y (IN REF. 4).
 FT CONFLICT 273 273 A -> P (IN REF. 4).
 FT CONFLICT 281 281 G -> A (IN REF. 4).
 FT CONFLICT 291 291 S -> R (IN REF. 4).
 FT CONFLICT 297 297 F -> S (IN REF. 2).
 FT CONFLICT 340 340 G -> A (IN REF. 1).
 FT CONFLICT 418 418 I -> V (IN REF. 4).
 FT CONFLICT 439 439 H -> Y (IN REF. 4).
 FT CONFLICT 459 459 K -> R (IN REF. 4).
 FT CONFLICT 514 514 A -> R (IN REF. 1).
 FT CONFLICT 632 632 H -> R (IN REF. 5).
 FT STRAND 25 29
 FT HELIX 32 45
 FT HELIX 46 48
 FT STRAND 53 57
 FT HELIX 61 69
 FT TURN 70 71
 FT STRAND 75 78
 FT HELIX 80 87
 FT TURN 89 91
 FT STRAND 93 102
 FT STRAND 107 108
 FT STRAND 110 118
 FT TURN 126 127
 FT TURN 129 130
 FT STRAND 133 135
 FT TURN 138 139
 FT TURN 141 144
 FT HELIX 145 155
 FT TURN 156 156
 FT TURN 159 161
 FT HELIX 164 169
 FT TURN 170 171
 FT STRAND 175 176
 FT TURN 178 179
 FT TURN 182 184
 FT HELIX 186 188
 FT TURN 190 191
 FT TURN 196 197
 FT TURN 199 200
 FT TURN 204 205
 FT TURN 207 208
 FT HELIX 210 219
 FT TURN 220 221
 FT STRAND 225 229

Query Match 100.0%; Score 47; DB 1; Length 708;
 Best Local Similarity 100.0%; Pred.No. 0.059;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKXRM 8
 |||||
 DB 20 APRKXRM 27

RESULT 2

TRFL_BUBBU STANDARD; PRT; 708 AA.
 ID TRFL_BUBBU
 AC 07698;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lactoferrin precursor (Lactoferrin).
 LTF.
 OS Bubalus bubalis (Domestic water buffalo).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bubalus.
 NCBI_Taxid=89462;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Paramasivam M., Thattaiyath B.D., Kumar A., Srinivasan A.,
 RA Singh T.P.;
 RT "cDNA sequence of Buffalo lactoferrin";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RX MEDLINE=20003130; Pubmed=10531476;
 RA Karthikeyan S., Paramasivam M., Yadav S., Srinivasan A., Singh T.P.;
 RT "Structure of buffalo lactoferrin at 2.5-A resolution using crystals
 grown at 303 K shows different orientations of the N and C lobes.";
 RL Acta Crystallogr. D 55:1805-1813 (1999).
 CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
 CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
 OF AN ANION, USUALLY BICARBONATE.
 CC -1- SUBUNIT: Monomer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AU005203; CA006441.1; -.
 DR PDB; 1CE2; 19-MAR-99.
 DR PDB; 1BY1; 13-JAN-99.
 DR InterPro; IPR001156; Transferrin.
 DR Pfam; PF00405; Transferrin; 2.
 DR PRINTS; PR00442; TRANSFERRIN.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 2.
 KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal; 3d-structure.
 FT SIGNAL 1 19
 FT CHAIN 20 708 LACTOTRANSFERRIN.
 FT DISULFID 28 64
 FT DISULFID 38 55
 FT DISULFID 134 217
 FT DISULFID 176 192
 FT DISULFID 179 202
 FT DISULFID 189 200

FT DISULFID 250 264
 FT DISULFID 367 399
 FT DISULFID 377 390
 FT DISULFID 424 703
 FT DISULFID 444 666
 FT DISULFID 476 551
 FT DISULFID 500 694
 FT DISULFID 510 524
 FT DISULFID 521 534
 FT DISULFID 592 606
 FT DISULFID 644 649
 FT METAL 79 79
 FT METAL 111 111
 FT METAL 211 211
 FT METAL 272 272
 FT METAL 414 414
 FT METAL 452 452
 FT METAL 545 545
 FT METAL 614 614
 FT BINDING 140 140
 FT BINDING 482 482
 FT CARBOHYD 282 282
 FT CARBOHYD 300 300
 FT CARBOHYD 495 495
 FT CARBOHYD 564 564
 FT TUR 21 22
 FT STRAND 25 29
 FT TUR 32 46
 FT HELIX 47 58
 FT STRAND 53 57
 FT TUR 61 69
 FT HELIX 70 71
 FT STRAND 75 78
 FT HELIX 80 87
 FT TUR 89 91
 FT STRAND 93 102
 FT STRAND 107 108
 FT STRAND 110 118
 FT TUR 119 120
 FT HELIX 125 127
 FT TUR 129 130
 FT STRAND 132 135
 FT TUR 138 139
 FT TUR 141 144
 FT HELIX 145 150
 FT TUR 151 151
 FT HELIX 152 155
 FT TUR 159 161
 FT HELIX 164 169
 FT TUR 170 171
 FT STRAND 174 176
 FT TUR 178 179
 FT TUR 182 184
 FT HELIX 186 189
 FT TUR 190 191
 FT HELIX 196 198
 FT TUR 199 199
 FT TUR 200 200
 FT TUR 204 205
 FT TUR 207 208
 FT HELIX 211 219
 FT TUR 220 221
 FT STRAND 225 229
 FT TUR 230 231
 FT HELIX 232 236
 FT HELIX 240 243
 FT TUR 244 245
 FT STRAND 246 249
 FT TUR 251 252
 FT STRAND 255 257
 FT HELIX 258 260
 FT TUR 261 263
 FT TUR 267 270
 FT STRAND 273 277

IRON 1.
 IRON 1.
 IRON 1.
 IRON 2.
 IRON 2.
 IRON 2.
 ANION (POTENTIAL).
 ANION (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT HELIX 283 297
FT TURN 299 300
FT TURN 307 308
FT TURN 312 313
FT TURN 321 322
FT STRAND 325 328
FT TURN 331 332
FT HELIX 335 351
FT HELIX 354 362
FT STRAND 364 369
FT HELIX 371 384
FT TURN 385 385
FT STRAND 388 393
FT HELIX 396 404
FT TURN 405 406
FT STRAND 410 413
FT HELIX 415 422
FT TURN 423 425
FT STRAND 426 434
FT TURN 438 441
FT TURN 444 446
FT STRAND 452 459
FT TURN 460 461
FT TURN 463 464
FT HELIX 467 469
FT TURN 471 472
FT STRAND 474 477
FT TURN 480 481
FT TURN 483 486
FT HELIX 487 497
FT TURN 503 505
FT STRAND 508 510
FT TURN 512 513
FT TURN 515 517
FT TURN 519 520
FT TURN 522 523
FT TURN 528 529
FT TURN 533 534
FT TURN 538 539
FT TURN 541 542
FT HELIX 544 553
FT TURN 554 555
FT STRAND 559 563
FT HELIX 564 569
FT TURN 570 570
FT HELIX 578 581
FT TURN 582 582
FT TURN 585 586
FT STRAND 588 591
FT TURN 593 594
FT STRAND 597 599
FT TURN 600 601
FT TURN 603 605
FT STRAND 609 611
FT STRAND 615 619
FT HELIX 620 637
FT TURN 639 640
FT TURN 642 647
FT TURN 650 651
FT TURN 654 655
FT TURN 662 663
FT STRAND 664 669
FT HELIX 676 680
FT HELIX 682 692
FT TURN 693 694
FT HELIX 698 705

```

Query Match 100.0%; Score 47; DB 1; Length 708;

Best Local Similarity 100.0%; Pred. No. 0.059; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 APRKNVRW 8

```

DB 20 APRKNVRW 27
RESULT 3
ID TRFL_CAPHI STANDARD; PRT; 708 AA.
AC Q29477; Q29479;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lactotransferrin precursor (Lactoferrin).
GN LTF.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OK NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Lee T., Yu S., Kim S., Lee K., Yu D.;
RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA le Provost F., Nocart M., Guerin G., Martin P.;
RT "Characterization of the goat lactoferrin cDNA. Assignment of the
RL Biochem. Biophys. Res. Commun. 203:1324-1332(1994).
CC -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC OF AN ANION, USUALLY BICARBONATE.
CC -1- SUBUNIT: Monomer (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U53857; AAA97958.1; -.
DR EMBL; X78902; CAAS5517.1; -.
DR HSSP; O77698; ICE2.
DR InterPro; IPR001156; Transferrin.
DR Pfam; PF00405; transferrin; 2.
DR PRINTS; PR00422; TRANSFERRIN.
DR SMART; SMO0094; TR_PFR; 2.
DR PROSITE; PS00205; TRANSFERRIN_1; 2.
DR PROSITE; PS00206; TRANSFERRIN_2; 2.
DR PROSITE; PS00207; TRANSFERRIN_3; 2.
KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 708 LACTOTRANSFERRIN.
FT REPEAT 20 363 1.
FT REPEAT 364 708 2.
FT DISULFID 28 64 BY SIMILARITY.
FT DISULFID 38 55 BY SIMILARITY.
FT DISULFID 134 217 BY SIMILARITY.
FT DISULFID 176 192 BY SIMILARITY.
FT DISULFID 189 200 BY SIMILARITY.
FT DISULFID 250 264 BY SIMILARITY.
FT DISULFID 367 399 BY SIMILARITY.
FT DISULFID 377 390 BY SIMILARITY.
FT DISULFID 424 703 BY SIMILARITY.
FT DISULFID 444 666 BY SIMILARITY.
FT DISULFID 476 551 BY SIMILARITY.

```

```

FT DISULFID 500 694 BY SIMILARITY.
FT DISULFID 510 524 BY SIMILARITY.
FT DISULFID 521 534 BY SIMILARITY.
FT DISULFID 592 606 BY SIMILARITY.
FT DISULFID 644 649 BY SIMILARITY.
FT METAL 79 79 IRON 1 (BY SIMILARITY).
FT METAL 111 111 IRON 1 (BY SIMILARITY).
FT METAL 211 211 IRON 1 (BY SIMILARITY).
FT METAL 272 272 IRON 1 (BY SIMILARITY).
FT METAL 414 414 IRON 2 (BY SIMILARITY).
FT METAL 452 452 IRON 2 (BY SIMILARITY).
FT METAL 545 545 IRON 2 (BY SIMILARITY).
FT METAL 614 614 IRON 2 (BY SIMILARITY).
FT BINDING 140 140 ANION (BY SIMILARITY).
FT BINDING 482 482 ANION (BY SIMILARITY).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 300 300 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 387 387 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 56 56 I -> V (IN REF. 2).
FT CONFLICT 88 88 L -> R (IN REF. 2).
FT CONFLICT 124 124 Q -> K (IN REF. 2).
FT CONFLICT 154 154 F -> P (IN REF. 2).
FT CONFLICT 304 304 S -> R (IN REF. 2).
FT CONFLICT 414 414 D -> G (IN REF. 2).
SQ SEQUENCE 708 AA; 77358 MW; F2EDA3C8359960D CRC64;

```

```

Query Match      100.0%; Score 47; DB 1; Length 708;
Best Local Similarity 100.0%; Pred. No. 0.059;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 APRKNVRW 8
Db 20 APRKNVRW 27

```

```

RESULT 4
TRFL HORSE STANDARD; PRT; 695 AA.
ID 07811;
AC 15-JUL-1999. (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lactotransferrin precursor (Lactoferrin) (Fragment).
GN LTF.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxId=9796;
RN [1]
RP SEQUENCE FROM N.A.
RA Paramasivam M., Srinivasan A., Yadav M.P., Singh T.P.;
RT "cDNA sequence of mare Lactoferrin."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RX MEDLINE=99296631; PubMed=10366507;
RC TISSUE=Milk;
RN X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RA Shatma A.K., Paramasivam M., Srinivasan A., Yadav M.P., Singh T.P.;
RT "Three-dimensional structure of mare deferent lactoferrin at 2.6-A
resolution."
RL J. Mol. Biol. 289:303-317(1999).
CC -I- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
OF AN ANION, USUALLY BICARBONATE.
CC -I- SUBUNIT: Monomer.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -I- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AJ010930; CAA09407.1; -.
CC PDB; 1BX; 02-DEC-98.
CC PDB; 1BTU; 02-FEB-99.
CC PDB; 1BTZ; 02-FEB-99.
CC PDB; 1P9B; 10-FEB-01.
CC PDB; 1I6B; 13-FEB-02.
CC PDB; 1QJM; 14-JAN-00.
CC InterPro; IPR001156; Transferrin.
CC Pfam; PF00405; transferrin; 2.
CC PRINTS; PR00422; TRANSFERRIN.
CC SMART; SM00094; TR_FER; 2.
CC PROSITE; PS00205; TRANSFERRIN 1; 2.
CC PROSITE; PS00206; TRANSFERRIN 2; 2.
CC PROSITE; PS00207; TRANSFERRIN 3; 1.
CC KW Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
CC Signal; 3D-structure.
CC FT NON_TER 1 1
CC FT SIGNAL <1 6
CC FT CHAIN 7 695 LACTOTRANSFERRIN.
CC FT REPEAT 351 695 1.
CC FT DISULFID 15 51 2.
CC FT DISULFID 25 42
CC FT DISULFID 121 204
CC FT DISULFID 163 179
CC FT DISULFID 166 189
CC FT DISULFID 176 187
CC FT DISULFID 237 251
CC FT DISULFID 354 386
CC FT DISULFID 364 377
CC FT DISULFID 411 690
CC FT DISULFID 431 653
CC FT DISULFID 463 538
CC FT DISULFID 487 681
CC FT DISULFID 497 511
CC FT DISULFID 508 521
CC FT DISULFID 579 593
CC FT DISULFID 631 636
CC FT METAL 66 66 IRON 1 (BY SIMILARITY).
CC FT METAL 98 98 IRON 1 (BY SIMILARITY).
CC FT METAL 198 198 IRON 1 (BY SIMILARITY).
CC FT METAL 259 259 IRON 1 (BY SIMILARITY).
CC FT METAL 401 401 IRON 2 (BY SIMILARITY).
CC FT METAL 439 439 IRON 2 (BY SIMILARITY).
CC FT METAL 532 532 IRON 2 (BY SIMILARITY).
CC FT METAL 601 601 IRON 2 (BY SIMILARITY).
CC FT BINDING 127 127 ANION (BY SIMILARITY).
CC FT BINDING 459 469 ANION (BY SIMILARITY).
CC FT CARBOHYD 133 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 287 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 482 482 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT STRAND 12 16
CC FT HELIX 19 34
CC FT TURN 35 36
CC FT STRAND 40 44
CC FT HELIX 48 56
CC FT TURN 57 58
CC FT STRAND 62 65
CC FT HELIX 67 74
CC FT TURN 76 78
CC FT STRAND 80 88
CC FT STRAND 95 95
CC FT STRAND 97 105
CC FT TURN 106 107
CC FT HELIX 112 114
CC FT TURN 116 117
CC FT STRAND 119 122

```

```

FT      TURN      125      126
FT      TURN      128      131
FT      HELIX      132      137
FT      HELIX      139      142
FT      TURN      147      148
FT      HELIX      151      158
FT      STRAND      161      163
FT      TURN      165      166
FT      TURN      169      171
FT      HELIX      173      175
FT      TURN      176      178
FT      TURN      183      187
FT      TURN      191      192
FT      TURN      194      195
FT      HELIX      197      206
FT      TURN      207      208
FT      STRAND      212      216
FT      HELIX      219      223
FT      TURN      227      230
FT      HELIX      231      232
FT      STRAND      233      237
FT      TURN      238      240
FT      STRAND      241      244
FT      HELIX      245      247
FT      TURN      248      250
FT      STRAND      254      257
FT      STRAND      260      264
FT      HELIX      270      284
FT      TURN      286      287
FT      TURN      294      295
FT      TURN      308      309
FT      STRAND      312      315
FT      HELIX      322      326
FT      HELIX      328      336
FT      TURN      337      338
FT      HELIX      341      349
FT      STRAND      351      356
FT      HELIX      358      370
FT      TURN      371      372
FT      STRAND      375      380
FT      HELIX      383      392
FT      TURN      393      393
FT      STRAND      397      400
FT      HELIX      402      410
FT      TURN      411      412
FT      STRAND      414      421
FT      HELIX      423      433
FT      STRAND      439      446
FT      HELIX      454      456
FT      TURN      458      459
FT      STRAND      462      464
FT      TURN      467      468
FT      TURN      470      473
FT      HELIX      474      484
FT      HELIX      489      491
FT      TURN      492      492
FT      STRAND      496      497
FT      TURN      499      500
FT      TURN      503      504
FT      TURN      506      507
FT      STRAND      508      508
FT      TURN      515      516
FT      TURN      520      521
FT      TURN      525      526
FT      TURN      528      529
FT      HELIX      531      540
FT      TURN      541      542
FT      STRAND      546      550
FT      HELIX      551      555
FT      TURN      556      557
FT      TURN      559      560
FT      HELIX      565      568
FT      TURN      569      569

```

```

FT      TURN      572      573
FT      STRAND      575      578
FT      TURN      580      581
FT      STRAND      584      586
FT      HELIX      587      592
FT      STRAND      596      598
FT      STRAND      602      605
FT      TURN      607      609
FT      HELIX      610      624
FT      TURN      626      627
FT      TURN      629      630
FT      HELIX      631      634
FT      TURN      637      638
FT      TURN      641      642
FT      TURN      649      650
FT      STRAND      651      655
FT      TURN      658      659
FT      HELIX      663      667
FT      HELIX      669      679
FT      TURN      680      681

Query Match      89.4%; Score 42; DB 1; Length 695;
Best Local Similarity 87.5%; Pred.No. 0.62;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      1 APRKNVW 8
DB      7 APRKSVW 14

RESULT 5
TRFL_PIG      STANDARD; PRT; 704 AA.
ID      P14632; Q29557;
AC      01-APR-1990 (rel. 14, Created)
DT      15-DEC-1998 (rel. 37, Last sequence update)
DT      28-FEB-2003 (rel. 41, Last annotation update)
DE      Lactoferrin precursor (Lactoferrin).
LTF.
OS      Sus scrofa (Pig).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX      NCBI_TaxID=9823;
RN      [1]
RP      MEDLINE=92379101; PubMed=1511016;
RX      Lyndon J.P., O'Malley B.R., Saucedo O., Lee T., Headon D.R.,
RA      Connely O.M.;
RT      "Nucleotide and primary amino acid sequence of porcine lactoferrin.";
RN      [2]
RP      MEDLINE=92379101; PubMed=1511016;
RX      Alexander L.J., Levine W.B., Teng C.T., Beattie C.W.;
RA      "Cloning and sequencing of the porcine lactoferrin cDNA.";
RT      Anim. Genet. 23:251-256(1992).
RN      [3]
RP      MEDLINE=9010538; PubMed=2605266;
RX      Hutcheon T.W., Magnuson J.S., Yip T.-T.;
RA      "Rapid purification of porcine colostrum whey lactoferrin by affinity
RT      chromatography on single-stranded DNA-agarose. Characterization,
RT      amino acid composition and N-terminal amino acid sequence.";
RN      [4]
RP      MEDLINE=999323329(1989);
RX      Biochim. Biophys. Acta 999:323-329(1989).
CC      -1- FUNCTION: TRANSFERRINS ARE IRON BINDING TRANSPORT PROTEINS WHICH
CC      CAN BIND TWO ATOMS OF FERRIC IRON IN ASSOCIATION WITH THE BINDING
CC      OF AN ANION, USUALLY BICARBONATE.
CC      -1- SUBUNIT: Monomer.
CC      -1- SUBCELLULAR LOCATION: Secreted.
CC      -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC      -1- SIMILARITY: BELONGS TO THE TRANSFERRIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M92089; AAA31102.1; -
 DR EMBL; M81327; AAA31059.1; -
 DR PIR; A45543; A45543.
 DR HSSP; 077698; 1CE2.
 DR InterPro: IPR001156; Transferrin.
 DR Pfam; PF00405; transferrin; 2.
 DR SMART; SM00094; TR_FER; 2.
 DR PROSITE; PS00205; TRANSFERRIN_1; 2.
 DR PROSITE; PS00206; TRANSFERRIN_2; 2.
 DR PROSITE; PS00207; TRANSFERRIN_3; 1.
 DR Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;
 KW Signal.
 FT CHAIN 1 19
 FT SIGNAL 704
 FT REPEAT 20 359
 FT REPEAT 360 704
 FT DISULFID 28 62
 FT DISULFID 38 53
 FT DISULFID 130 213
 FT DISULFID 172 188
 FT DISULFID 185 196
 FT DISULFID 246 260
 FT DISULFID 363 395
 FT DISULFID 373 386
 FT DISULFID 420 699
 FT DISULFID 472 547
 FT DISULFID 496 630
 FT DISULFID 506 520
 FT DISULFID 517 530
 FT DISULFID 588 602
 FT DISULFID 640 645
 FT METAL 77 77
 FT METAL 107 107
 FT METAL 207 207
 FT METAL 268 268
 FT METAL 410 410
 FT METAL 448 448
 FT METAL 541 541
 FT METAL 610 610
 FT BINDING 136 136
 FT BINDING 478 478
 FT CARBOHYD 385 385
 FT CARBOHYD 491 491
 FT CONFLICT 12 12
 FT CONFLICT 46 48
 FT CONFLICT 51 51
 FT CONFLICT 85 85
 FT CONFLICT 121 121
 FT CONFLICT 132 132
 FT CONFLICT 284 284
 FT CONFLICT 573 573
 FT CONFLICT 590 590
 FT CONFLICT 625 625
 FT CONFLICT 662 662
 FT CONFLICT 704 704
 SQ SEQUENCE 704 AA; 77625 MW; 93261EPD608AD58 CRC64;
 Query Match 80.9%; Score 38; DB 1; Length 704;
 Best Local Similarity 75.0%; Pred. No. 4.2;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKQVRW 8
 ||:|:|:|
 DB 20 APRKQVRW 27

RESULT 6
 ID POINT SINDV STANDARD; PRT; 2512 AA.
 AC P03317;
 DT 21-0UL-1986 (Rel. 01, Created)
 DT 21-0UL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nonstructural polyprotein (p270) [Contains: Nonstructural protein
 DE NSP1; Nonstructural NSP2; Nonstructural NSP3; Nonstructural NSP4].
 OS Sindbis virus (strain HSSP).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphaviruses.
 OX NCB1:Taxid=11034;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84148439; PubMed=6322438;
 RA Straus E.G., Rice C.M., Straus J.H.;
 RT "Complete nucleotide sequence of the genomic RNA of Sindbis virus,";
 RL Virology 133:92-110(1984).
 RN [2]
 RP SEQUENCE OF 1-54 FROM N.A.
 RX MEDLINE=83268700; PubMed=6308269;
 RA Ou J.H., Straus E.G., Straus J.H.;
 RT "The 5'-terminal sequences of the genomic RNAs of several
 RT alphaviruses,";
 RL J. Mol. Biol. 168:1-15(1983).
 RN [3]
 RP SEQUENCE OF 1429-2512 FROM N.A.
 RX MEDLINE=83299955; PubMed=6577423;
 RA Straus E.G., Rice C.M., Straus J.H.;
 RT "Sequence coding for the alphavirus nonstructural proteins is
 RT interrupted by an opal termination codon,";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:5271-5275(1983).
 RN [4]
 RP SEQUENCE OF 2431-2512 FROM N.A.
 RX MEDLINE=83039346; PubMed=6291034;
 RA Ou J.H., Rice C.M., Dalgarno L., Straus E.G., Straus J.H.;
 RT "Sequence studies of several alphavirus genomic RNAs in the region
 RT containing the start of the subgenomic RNA,";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:5235-5239(1982).
 CC -1- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
 CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: RADYDROUGH OF THE TERMINATOR CODON UGA OCCURS
 CC BETWEEN THE CODONS FOR 1896-TYR AND 1897-LEU.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; J02363; AAA96975.1; ALT_SEQ.
 DR PIR; A03917; MNWVS.
 DR MEROPS; C09_001; -
 DR InterPro; IPR002589; Alpp.
 DR InterPro; IPR002620; Peptidase C9.
 DR InterPro; IPR001788; RNA_dep_RNApol12.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR InterPro; IPR006066; Viral_helicase1.
 DR Pfam; PF01661; Alpp; 1.
 DR Pfam; PF01707; Peptidase_C9; 1.
 DR Pfam; PF00978; RNA_dep_RNApol12; 1.
 DR Pfam; PF01443; Viral_helicase1; 1.
 DR SMART; SM00506; Alpp; 1.
 KW Polyprotein; Nonstructural protein; RNA-binding; Helicase.
 FT CHAIN 1 540
 FT CHAIN 541 1347
 FT CHAIN 1348 1896
 FT CHAIN 1897 2512
 NONSTRUCTURAL PROTEIN NSP1.
 NONSTRUCTURAL PROTEIN NSP2.
 NONSTRUCTURAL PROTEIN NSP3.
 NONSTRUCTURAL PROTEIN NSP4.

SEQ SEQUENCE 2512 AA; 279546 MW; F3656FC8BB495726 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 2512;

Best Local Similarity 62.5%; Pred. No. 64;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 APRKNVRW 8

DB 1195 APRKRIEW 1202

RESULT 7

VCOM ADEB12 STANDARD; PRT; 347 AA.

AC P36717;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DE Minor core protein (Protein V).

GN PV.

OS Human adenovirus type 12.

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.

OX NCBI_TaxID=28282;

RP SEQUENCE FROM N.A.

RX MEDLINE=94076430; PubMed=8254750;

RA Sprengel J., Schmitz B., Heuss-Neitzel D., Zock C., Doerfler W.;

RT "Nucleotide sequence of human adenovirus type 12 DNA: comparative functional analysis."

RL J. Virol. 68:379-389(1994).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X73487; CA51888.1; -

DR PIR; S33939; S33939.

DR InterPro; IPR005608; Adeno_PV.

DR Pfam; PF03910; Adeno_PV; 1.

KW Core protein; Late protein.

SEQUENCE 347 AA; 39369 MW; 3C338F62A5A53027 CRC64;

Query Match 72.3%; Score 34; DB 1; Length 347;

Best Local Similarity 62.5%; Pred. No. 13;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8

DB 61 APRRRVOW 68

RESULT 8

VCOM ADEA40 STANDARD; PRT; 358 AA.

AC P48753;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DE Minor core protein (Protein V).

GN PV.

OS Human adenovirus type 40.

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.

OX NCBI_TaxID=28284;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94087748; PubMed=8263936;

RA Davison A.J., Telford E.A., Watson M.S., McBride K., Mautner V.;

RT "The DNA sequence of adenovirus type 40.";

RL J. Mol. Biol. 234:1308-1316(1993).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L19443; AAC13964.1; -

DR InterPro; IPR005608; Adeno_PV.

DR Pfam; PF03910; Adeno_PV; 1.

KW Core protein; Late protein.

SEQUENCE 358 AA; 40707 MW; B7D0D3BAF794C96E CRC64;

Query Match 72.3%; Score 34; DB 1; Length 358;

Best Local Similarity 62.5%; Pred. No. 14;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8

DB 63 APRRRVOW 70

RESULT 9

VCOM ADE05 STANDARD; PRT; 368 AA.

AC P24938;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DE Minor core protein (Protein V).

GN PV.

OS Human adenovirus type 5.

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.

OX NCBI_TaxID=28285;

RP SEQUENCE FROM N.A.

RX MEDLINE=92087470; PubMed=1727603;

RA Chodoczek J., Bieder F., Jactot B.;

RT "The sequence of the genome of adenovirus type 5 and its comparison

RT with the genome of adenovirus type 2.";

RL Virology 186:280-285(1992).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M73260; AAA96409.1; -

DR PIR; C39449; FOADM5.

DR InterPro; IPR005608; Adeno_PV.

DR Pfam; PF03910; Adeno_PV; 1.

KW Core protein; Late protein.

SEQUENCE 368 AA; 41446 MW; 722E6C6D22C692A4 CRC64;

Query Match 72.3%; Score 34; DB 1; Length 368;

Best Local Similarity 62.5%; Pred. No. 14;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8

DB 59 APRRRVOW 66

RESULT 10

VCOM ADE02 STANDARD; PRT; 369 AA.

AC P03267;

DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Minor core protein (Protein V).
 GN PV.
 OS Human adenovirus type 2.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OC NCBI_TaxID=10515;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=65054835; PubMed=6094534;
 RA Alestrom P., Akusjall G., Lager M., Yeh-Kai L., Petersson U.,
 RT "Genes encoding the core proteins of adenovirus type 2.";
 RL J. Biol. Chem. 259:13980-13985(1984).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J01917; AAA92213.1; -
 DR PIR: A03837; FOADM2.
 DR InterPro: IPR005608; Adeno_PV.
 DR Pfam: PF03910; Adeno_PV; 1-
 KW CORE protein; late protein.
 SQ SEQUENCE 369 AA; 41721 MW; 577C9E645EB5E7DE CRC64;
 Qy Query Match 72.3%; Score 34; DB 1; Length 369;
 Best Local Similarity 62.5%; Pred. No. 14;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Db 1 APPRISVW 8
 60 AARRRVQW 67
 RESULT 11
 RT30 HUMAN STANDARD; PRT; 439 AA.
 AC Q9NP92; Q96191; Q9H0P8; Q9NSP9; Q9NZ76;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Mitochondrial 28S ribosomal protein S30 (S30mt) (MNP-S30) (Programmed
 DE cell death protein 9) (BM047).
 GN MRPS30 OR PDCD9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
 RX MEDLINE=20108792; PubMed=10640817;
 RA Carim L., Sumoy L., Nadal M., Estivill X., Baccarcel M.,
 RT "Cloning, expression, and mapping of PDCD9, the human homolog of
 RT Gallus gallus pro-apoptotic protein p21.";
 RL Cytogenet. Cell Genet. 87:85-88(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Carim L., Estivill X., Baccarcel M., Sumoy L.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=21154917; PubMed=11230166;
 RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glasel S.,
 RA Anorge W., Boecker M., Bloeker H., Baetsachs S., Blum H.,
 RA lauber J., Duesterhoeft A., Beyer A., Koerner K., Strack N.,
 RA Mews H.-W., Ottenwälder B., Obernauer B., Tampe J., Heubner D.,
 RA Wambutt R., Korn B., Klein M., Poustka A.;

RT "Towards a catalog of human genes and proteins: sequencing and
 RT analysis of 500 novel complete protein coding human cDNAs.";
 RL Genome Res. 11:422-435(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=reticulocytoma;
 RA Isogaki T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
 RA Suzuki Y., Nagai K., Sugano S., Iehi S., Kawai-Hito Y., Saito K.,
 RA Yamamoto Y., Makamatsu A., Nakamura Y., Kojima S., Nagahara K.,
 RA Masuno Y., Ota T., Okano K., Yoshikawa Y., Kotuka S., Sasaki N.,
 RA Hattori A., Okumura K., Iwayanagi T., Nishimura K.,
 RT "NEO human cDNA sequencing project.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stalcenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uebin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muliyil S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Villalón D.K., Mizny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Faney J., Helton B., Kettelman M., Madan A., Kodriges S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.W., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [6]
 RP SEQUENCE OF 89-439 FROM N.A.
 RC TISSUE=Bone marrow;
 RA Zhao M., Gu J., Li N., Peng Y., Han Z., Chen Z.;
 RT "A novel gene expressed in human bone marrow.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP IDENTIFICATION.
 RX MEDLINE=21276436; PubMed=11279123;
 RA Koc E.C., Burkhart W., Blackburn K., Moseley A., Spremulli L.L.;
 RT "The small subunit of the mammalian mitochondrial ribosome:
 RT identification of the full complement of ribosomal proteins present.";
 RL J. Biol. Chem. 276:19363-19374(2001).
 CC -1- SUBUNIT: Component of the mitochondrial ribosome small subunit
 CC (28S) which comprises a 12S rRNA and about 30 distinct proteins.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.
 CC -1- TISSUE SPECIFICITY: Heart, skeletal muscle, kidney and liver.
 CC Lower expression in placenta and peripheral blood leukocytes.
 CC -1- CAUTION: Ref.6 sequence differs from that shown due to a number of
 CC sequencing errors that include framehifts.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF146192; AAF65227.1; -
 DR EMBL: AL355715; CAB90810.1; -
 DR EMBL: AL355715; CAB90811.1; -
 DR EMBL: AL35706; CAB66641.1; -
 DR EMBL: AK074777; BAC11202.1; -
 DR EMBL: BC007735; AAH07735.1; -
 DR EMBL: AF217523; AAF67634.1; ALT_FRAME.

DR Genew; HGNC:8769; MRP830.
 DR GO; GO:0006915; P:apoptosis; TAS.
 KW Ribosomal protein; Mitochondrion.
 FT CONFLICT 33 33 S -> C (IN REF. 4 AND 5).
 FT CONFLICT 82 82 I -> V (IN REF. 3).
 SQ SEQUENCE 439 AA; 50348 MM; AD0564743E03290C CRC64;
 Query Match 72.3%; Score 34; DB 1; Length 439;
 Best Local Similarity 71.4%; Pred. No. 17;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 PRONRW 8
 DB 388 PRKNICW 394
 RESULT 12
 LITA HUMAN STANDARD; PRT; 166 AA.
 ID LITA_HUMAN
 AC P05451; P11379;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-OCT-1989 (Rel. 12, Last annotation update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Lithostathine I alpha precursor (Pancreatic stone protein) (PSP)
 DE (Pancreatic thread protein) (PTP) (Islet of Langerhans regenerating
 DE protein) (REG) (Regenerating protein I alpha) (Islet cells
 DE regeneration factor) (ICRF).
 GN REGIA OR REG OR PSPS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88115343; PubMed=2963000;
 RA Terazono K., Yamamoto H., Takasawa S., Shiga K., Yonemura Y.,
 RA Tochino Y., Okamoto H.;
 RT "A novel gene activated in regenerating islets.";
 RL J. Biol. Chem. 263:2111-2114(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90237042; PubMed=2332435;
 RA Watanabe T., Yonekura H., Terazono K., Yamamoto H., Okamoto H.;
 RT "Complete nucleotide sequence of human reg gene and its expression in
 RT normal and tumoral tissues. The reg protein, pancreatic stone
 RT protein, and pancreatic thread protein are one and the same product
 RT of the gene.";
 RL J. Biol. Chem. 265:7432-7439(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=89292148; PubMed=2525567;
 RA Giorgi D., Bernard J.-P., Rouquier S., Iovanna J., Sables H.,
 RA Dagnon J.-C.;
 RT "Secretory pancreatic stone protein messenger RNA. Nucleotide
 RT sequence and expression in chronic calcifying pancreatitis.";
 RL J. Clin. Invest. 84:100-106(1989).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Boonyasrisawat W., Tandhanand-Banchuin N., Vannasaeng S.,
 RA Yenchitsomanus P.;
 RT Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altmann S.F., Zeeberg B., Buertow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherob A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [6]
 RP SEQUENCE OF 34-166.
 RX MEDLINE=88029417; PubMed=3565916;
 RA de Caro A.M., Bonicel J.J., Rouimi P., de Caro J.D., Sables H.,
 RA Rovey M.;
 RT "Complete amino acid sequence of an immunoreactive form of human
 RT pancreatic stone protein isolated from pancreatic juice.";
 RL Eur. J. Biochem. 168:201-207(1987).
 RN [7]
 RP SEQUENCE OF 34-98.
 RX MEDLINE=8709950; PubMed=3541906;
 RA Montalco G., Bonicel J.J., Multigner L., Rovey M., Sables H.,
 RA de Caro A.M.;
 RT "Partial amino acid sequence of human pancreatic stone protein, a
 RT novel pancreatic secretory protein.";
 RL Biochem. J. 238:227-232(1986).
 RN [8]
 RP SEQUENCE OF 34-78.
 RX MEDLINE=86086356; PubMed=3908481;
 RA Gross J.R., Carlson R.I., Brauer A.W., Margolies M.N., Marshaw A.L.,
 RA Wands J.R.;
 RT "Isolation, characterization, and distribution of an unusual
 RT pancreatic human secretory protein.";
 RL J. Clin. Invest. 76:2115-2125(1985).
 RN [9]
 RP SEQUENCE OF 23-47.
 RX MEDLINE=89150292; PubMed=2493268;
 RA de Caro A.M., Adrich Z., Fournet B., Capon C., Bonicel J.J.,
 RA de Caro J.D., Rovey M.;
 RT "N-terminal sequence extension in the glycosylated forms of human
 RT pancreatic stone protein. The 5-oxoprolinyl N-terminal chain is O-
 RT glycosylated on the 5th amino acid residue.";
 RL Biochim. Biophys. Acta 994:281-284(1989).
 RN [10]
 RP SEQUENCE OF 33-58.
 RX MEDLINE=87219142; PubMed=3108036;
 RA Rouimi P., Bonicel J.J., Rovey M., de Caro A.;
 RT "Cleavage of the Arg-Ile bond in the native polypeptide chain of
 RT human pancreatic stone protein.";
 RL FEBS Lett. 216:195-199(1987).
 RN [11]
 RP IDENTITY OF REG WITH PSP.
 RX MEDLINE=89350859; PubMed=2764894;
 RA Stewart T.A.;
 RT "The human reg gene encodes pancreatic stone protein.";
 RL Biochem. J. 260:622-623(1989).
 RN [12]
 RP DISULFIDE BONDS.
 RX MEDLINE=91032149; PubMed=2226837;
 RA Itoh T., Tezuka H., Karoh T., Terazono K., Matsumoto K., Yoshida N.,
 RA Terazono K., Watanabe T., Yonekura H., Yamamoto H., Okamoto H.;
 RT "Isolation and characterization of human reg protein produced in
 RT Saccharomyces cerevisiae.";
 RL FEBS Lett. 272:85-88(1990).
 RN [13]
 RP ALZHEIMER'S DISEASE AND DEVELOPMENTAL EXPRESSION REGULATION.
 RX MEDLINE=90368981; PubMed=2394826;
 RA de la Monte S.M., Ozuturk M., Wands J.R.;
 RT "Enhanced expression of an exocrine pancreatic protein in Alzheimer's
 RT disease and the developing human brain.";

RL J. Clin. Invest. 86:1004-1013(1990).
 RN [14]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
 RX MEDLINE:96256285; PubMed:8654365;
 RA Bertrand J.A., Pignol D., Bernard J.-P., Verdier J.-M., Dagorn J.-C.,
 RA Fontecilla-Camps J.C.,
 RT "Crystal structure of human lithostathine, the pancreatic inhibitor
 RT of stone formation."
 RL EMBO J. 15:2678-2684(1996).
 RN [15]
 RP X-RAY CRYSTALLOGRAPHY (1.30 ANGSTROMS) OF 23-166.
 RX MEDLINE:20092874; PubMed:10625646;
 RA Gertraud V., Pignol D., Lore E., Bertrand J.A., Berland Y.,
 RA Fontecilla-Camps J.C., Canselier J.P., Gabas N., Verdier J.M.,
 RT "Mechanism of calcite crystal growth inhibition by the N-terminal
 RT undcapeptide of lithostathine."
 RL J. Biol. Chem. 275:1057-1064(2000).
 RN [16]
 RP STRUCTURE BY NMR OF 34-164.
 RX MEDLINE:97120677; PubMed:8961348;
 RA Patard L., Stoven V., Gharib B., Bontems F., Lallemand J.-Y.,
 RA de Reggi M.,
 RT "What function for human lithostathine? structural investigations by
 RT three-dimensional structure modelling and high-resolution NMR
 RT spectroscopy."
 RL Protein Eng. 9:949-957(1996).
 CC -1- FUNCTION: MIGHT ACT AS AN INHIBITOR OF SPONTANEOUS CALCIUM
 CC CARBONATE PRECIPITATION. MAY BE ASSOCIATED WITH NEURONAL
 CC SPROUTING IN BRAIN, AND WITH BRAIN AND PANCREAS REGENERATION.
 CC -1- TISSUE SPECIFICITY: IN PANCREATIC ACINAR CELLS AND IN LOWER
 CC LEVELS, IN BRAIN.
 CC -1- DEVELOPMENTAL STAGE: HIGH EXPRESSION LEVELS IN FETAL AND INFANT
 CC BRAINS; MUCH LOWER IN ADULT BRAINS.
 CC -1- DISEASE: ALZHEIMER'S DISEASE AND DOWN'S SYNDROME PATIENTS SHOW
 CC ENHANCED EXPRESSION OF PSP-RELATED TRANSCRIPTS AND INTRANEURONAL
 CC ACCUMULATION OF PSP-LIKE PROTEINS IN THEIR BRAINS.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M27180; AAA60546.1; -
 DR EMBL: M27189; AAA60545.1; -
 DR EMBL: M18963; AAA6558.1; -
 DR EMBL: J05412; AAA6559.1; -
 DR EMBL: AF172331; AAD51330.1; -
 DR EMBL: BC005350; AAH05350.1; -
 DR PIR: A35197; RGH01A.
 DR PIR: A45751; A45751.
 DR PDB: 1LIT; 11-JAN-97.
 DR PDB: 1ODD; 24-JAN-01.
 DR Genew: HGNC:19951; REG1A.
 DR MIM: 167770; -
 DR MIM: 167800; -
 DR GO: GO:0008284; P:positive regulation of cell proliferation; TAS.
 DR InterPro: IPR002353; Antifreeze1.
 DR InterPro: IPR001304; Lectin_C.
 DR InterPro: IPR003990; Pancreatic_ac.
 DR Pfam: PF00059; Lectin_c; 1.
 DR PRINTS: PRO1504; NCRCBATT1AP.
 DR PRINTS: PRO0356; ANTI-FREEZE1.
 DR SMART: SM00034; CLECT; 1.
 DR PROSITE: PS00615; C TYPE LECTIN 1; 1.
 DR PROSITE: PS00411; C TYPE LECTIN 2; 1.
 KW Glycoprotein; Signal; Alzheimer's disease; Down's syndrome; Lectin;
 KW 3D-structure; Pyroliidone carboxylic acid.
 FT SIGNAL 1 22

Query Match 70.2%; Score 33; DB 1; Length 166;
 Best Local Similarity 71.4%; Pred. No. 10;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 PRKNRW 8
 DB 105 PRKNRW 111
 RESULT 13
 ID LITB_HUMAN STANDARD; PRT; 166 AA.
 ID LITB_HUMAN
 ID P48304;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Lithostathine 1 beta precursor (Regenerating protein 1 beta).
 GN REG1B OR REG1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE:9415397; PubMed=8110835;
 RA Morizumi S., Watanabe T., Umno M., Nakagawara K.I., Suzuki Y.,
 RA Miyashita H., Yonekura H., Okamoto H.,
 RT "Isolation, structural determination and expression of a novel reg
 RT gene, human reg1 beta."
 RL Biochim. Biophys. Acta 1217:199-202(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:93351647; PubMed=8348956;
 RA Barroil C., Gharib B., Giorgi D., Sansonetti A., Dagorn J.-C.,
 RA Berge-Leffanc U.,
 RT "A gene homologous to the reg gene is expressed in the human
 RT pancreas."
 RL FEBS Lett. 327:289-293(1993).
 RN [3]
 RP CARBOHYDRATE-LINKAGE SITE.
 RX MEDLINE:9531286; PubMed=7607222;
 RA De Reggi M., Capon C., Gharib B., Wieruszski J.M., Michel R.,
 RA Fournet B.,
 RT "The glycane moiety of human pancreatic lithostathine. Structure
 RT characterization and possible pathophysiological implications."
 RL Eur. J. Biochem. 230:503-510(1995).
 CC -1- FUNCTION: MIGHT ACT AS AN INHIBITOR OF SPONTANEOUS CALCIUM
 CC CARBONATE PRECIPITATION. MAY BE ASSOCIATED WITH NEURONAL
 CC SPROUTING IN BRAIN, AND WITH BRAIN AND PANCREAS REGENERATION.
 CC -1- PTM: ALL O-LINKED GLYCANS CONSIST OF GAL-GLCNAc-GAL-GALNAc
 CC TETRASACCHARIDE CORE AND GET ELONGATED (MICROHETEROGENEITY).
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D17291; BA004124.1; -
 DR EMBL: D16816; BA004091.1; -
 DR EMBL: L08010; AAA18204.1; -
 DR PIR: S34591; RGH01B.
 DR HSSP: P05451; 1LIT.
 DR Genew: HGNC:19952; REG1B.
 DR MIM: 167771; -
 DR GO: GO:0008283; P:cell proliferation; TAS.
 DR InterPro: IPR001304; Lectin_C.
 DR InterPro: IPR003990; Pancreatic_ac.
 DR Pfam: PF00059; Lectin_c; 1.

```

DR PRINTS; PRO1504; ENCEPATITAP.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C TYPE LECTIN 1; 1.
DR PROSITE; PS00641; C TYPE LECTIN 2; 1.
DR GLYCOPROTEIN; Signal; Lectin; Pyroglutamate carboxylic acid.
FT SIGNAL 1 22
FT CHAIN 23 166
FT DOMAIN 34 164
FT MOD_RES 23 23
FT CARBOHYD 27 27
FT DISULFID 36 47
FT DISULFID 64 162
FT DISULFID 137 154
SQ SEQUENCE 166 AA; 18665 MW; 11DC20E11A5DDE8 CRC64;

Query Match
Best Local Similarity 70.2%; Score 33; DB 1; Length 166;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRKNVW 8
Db 105 PKKNRW 111

RESULT 14
R15E ARCFU STANDARD; PRT; 194 AA.
AC 02795;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L15e.
GN RPL15E OR AP2319.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kierulff A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Kachuravsky J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Ullrich T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -1- SIMILARITY: BELONGS TO THE L15E FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; AE000944; AA88937.1; -.
CC PIR; G69539; G69539.
CC TIGR; AE2319; -.
CC HAMAP; MF_00256; -; 1.
DR InterPro; IPR000439; Ribosomal_L15e.
DR Pfam; PF00827; Ribosomal_L15e; 1.
DR PROSITE; PS01194; RIBOSOMAL_L15E; 1.

```

```

KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 194 AA; 23167 MW; 4B607B60F132BD23 CRC64;

Query Match
Best Local Similarity 70.2%; Score 33; DB 1; Length 194;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRKNVW 8
Db 93 PKKNRW 99

RESULT 15
BTUC_SALTI STANDARD; PRT; 326 AA.
ID BTUC_SALTI
AC 082615;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Vitamin B12 transport system permease protein btuc.
DE BTUC OR STY1770 OR T1221.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parthill C., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Hague A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skellon J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=2253167; PubMed=12644504;
RA Deng M., Itoh S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burdand V., Kodymiani V., Schwartz D.C., Blatter F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- FUNCTION: Part of the binding-protein-dependent transport system
CC for vitamin B12. Involved in the translocation of the substrate
CC across the membrane (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM PERMEASE FAMILY. FECCD SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; AL627271; CAD02012.1; -.
CC PIR; AE016838; AA068876.1; -.
CC HAMAP; MF_01004; -; 1.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR000522; FECCD.
DR Pfam; PF01032; FECCD; 1.
DR ProDom; PD001557; FECCD; 1.
DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; FALSE_NEG.

```

KW Transport; Cobalt transport; Transmembrane; Inner membrane;
KM Complete proteome.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 113 123 POTENTIAL.
FT TRANSMEM 146 166 POTENTIAL.
FT TRANSMEM 164 204 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 274 294 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
SQ SEQUENCE 326 AA; 34864 MW; 3835B836331225CB CRC64;

Query Match 70.2%; Score 33; DB 1; Length 326;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RKNVRW 8
DB 10 RKNVRW 15

Search completed: July 30, 2003, 16:24:33
Job time : 3.61905 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2003, 16:16:55 ; Search time 5.09524 Seconds
(without alignments)
150.994 Million cell updates/sec

Title: US-09-787-070-8

Perfect score: 47

Sequence: 1 APRKNVW 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_76: *
2: pirl: *
3: pirl2: *
4: pirl3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	100.0	33	2 S52107	lactoferrin - sheep
2	47	100.0	708	1 TFBOL	lactotransferrin p
3	47	100.0	708	2 JC2323	lactoferrin - goat
4	38	80.9	703	2 A45543	lactoferrin precu
5	35	74.5	345	2 T25580	hypothetical prote
6	35	74.5	379	2 AD3390	phm protein (impo
7	35	74.5	404	2 B89819	pyrimidine nucleos
8	35	74.5	2512	1 MNWYS	nonstructural poly
9	34	72.3	347	2 S33939	minor core protein
10	34	72.3	368	1 FOADM5	minor core protein
11	34	72.3	369	1 FOADM2	minor core protein
12	33	70.2	85	2 T17882	hypothetical prote
13	33	70.2	164	2 T30991	hypothetical prote
14	33	70.2	166	1 RGRH1A	hypothetical prote
15	33	70.2	166	1 RGRH1B	regenerating islet
16	33	70.2	166	1 A45751	pancreatic stone p
17	33	70.2	194	2 G69539	ribosomal protein
18	33	70.2	227	2 S72574	hypothetical prote
19	33	70.2	253	2 T30381	hypothetical prote
20	33	70.2	326	2 AC0705	vitamin B12 transp
21	33	70.2	466	2 T30040	hypothetical prote
22	33	70.2	468	2 S46791	hypothetical prote
23	33	70.2	489	2 T13026	hypothetical prote
24	33	70.2	694	2 B54901	long-chain-fatty-a
25	33	70.2	776	2 T29064	hyaluronate lyase
26	33	70.2	1451	2 S65571	pattern formation
27	33	70.2	1795	2 T30332	avirlence protein
28	33	70.2	3759	2 A35085	criticorx protein
29	33	68.1	56	2 F82830	hypothetical prote

30	32	68.1	96	2 A87110	conserved hypothet
31	32	68.1	153	2 S77791	hypothetical prote
32	32	68.1	245	2 T26868	hypothetical prote
33	32	68.1	245	2 G70985	hypothetical prote
34	32	68.1	276	2 T50896	hypothetical membr
35	32	68.1	330	2 A87194	polyprenyl diphosp
36	32	68.1	367	2 H86425	hypothetical prote
37	32	68.1	379	2 F69816	reticuline oxidase
38	32	68.1	404	2 B95946	probable enzyme pr
39	32	68.1	476	2 A69467	hypothetical prote
40	32	68.1	508	2 D71264	conserved hypothet
41	32	68.1	508	2 C83138	conserved hypothet
42	32	68.1	661	2 C83843	hypothetical prote
43	32	68.1	694	2 S56060	long-chain-fatty-a
44	32	68.1	741	2 S50340	CH12 protein - Ye
45	32	68.1	771	2 T34376	hypothetical prote

ALIGNMENTS

RESULT 1

S52107
Lactoferrin - sheep (fragment)
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 14-Jul-1995 #sequence_revision 05-Dec-1997 #text_change 05-Dec-1997
C:Accession: S52107
R:Qian, Z.Y.; Jollies, P.; Migliore-Samour, D.; Fiat, A.M.
Biochim. Biophys. Acta 1243, 25-32, 1995
A:Title: Isolation and characterization of sheep lactoferrin, an inhibitor of platelet a
A:Reference number: S52107; MUID:95127729; PMID:7827104
A:Accession: S52107
A>Status: preliminary
A:Molecule type: protein
A:Residues: 1-33 <Q1A>
C:Superfamily: transferrin; transferrin repeat homology
C:Keywords: duplication

Query Match
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVW 8

DB 1 APRKNVW 8

RESULT 2

TFBOL
Lactotransferrin precursor - bovine

N:Alternate names: lactoferrin
C:Species: Bos primigenius taurus (cattle)
C:Date: 31-Mar-1992 #sequence_revision 21-Nov-1997 #text_change 11-May-2000
C:Accession: I45919; S14674; S14110; S18517; JTO595; S13097; S18518; S13881; PLO148; S21

R:Tsang, T.C.; Burns, D.K.; Wang, F.; Pan, Y.
FASEB J. 6, 233, 1991
A:Title: Cloning of a 80-kD advanced glycosylation end product (AGE) binding protein fro

A:Reference number: I45919
A:Accession: I45919

A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A:Residues: 1-708 <TSA>
A:Cross-references: GB:I08604; NID:9163269; PIDN:AAA30609.1; PID:9163270

R:Pierce, A.
submitted to the EMBL Data Library, November 1990

A:Reference number: S14674
A:Accession: S14674

A:Molecule type: mRNA

A:Residues: 1-144, 'V', 146-163, 'PP', 166-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PII>

A:Cross-references: EMBL:X57084; NID:9505; PIDN:CAA40366.1; PID:9506

R:Pierce, A.; Colavizza, D.; Benalissa, M.; Maes, P.; Tartat, A.; Montreuil, J.; Spik, G
Eur. J. Biochem. 196, 177-184, 1991

A:Title: Molecular cloning and sequence analysis of bovine lactotransferrin.

A:Reference number: S14110; MUID:9160550; PMID:2001696
 A:Accession: S14110
 A:Molecule type: mRNA
 A:Residues: 3-144; 'V', 146-339, 'A', 341-438, 'Y', 440-513, 'R', 515-708 <PI2>
 A:Cross-references: EMBL:X57064
 A:Accession: S18517
 A:Molecule type: protein
 A:Residues: 20-35; 82-114; 148-163, 'P', 166-178, 'V', 'P', 183-190; 205-212; 230-239; 304-339; 559
 R:Goodman, R.E.; Schanbacher, F.L.
 Biochem. Biophys. Res. Commun. 180, 75-84, 1991
 A:Title: Bovine lactoferrin mRNA: sequence, analysis, and expression in the mammary gland
 A:Reference number: U70595; MUID:92028986; PMID:1718281
 A:Accession: U70595
 A:Molecule type: mRNA
 A:Residues: 1-65, 'PG', 68-296, 'S', 298-339, 'A', 341-708 <GOO>
 A:Cross-references: GB:M63502
 A:Note: the authors translated the codon CCG for residue 66 as Arg and TCT for residue 2
 R:Mead, P.B.; Tweedie, J.W.
 Nucleic Acids Res. 18, 7167, 1990
 A:Title: cDNA and protein sequence of bovine lactoferrin.
 A:Reference number: S13097; MUID:91088328; PMID:2263492
 A:Accession: S13097
 A:Molecule type: mRNA
 A:Residues: 28-33, 'DS', 36-38, 'P', 40-708 <MEA>
 A:Cross-references: EMBL:X54801
 A:Accession: S18518
 A:Molecule type: protein
 A:Residues: 20-47; 59-66; 132-139; 256-277; 278, 305-332; 343-351; 361-363; 586, 587-589; 598-619
 R:Mead, P.B.
 submitted to the EMBL Data Library, October 1990
 A:Reference number: S13881
 A:Accession: S13881
 A:Molecule type: mRNA
 A:Residues: 28-38, 'P', 40-86, 'C', 88-708 <ME3>
 A:Cross-references: EMBL:X54801
 R:Rejman, J.V.; Hegarty, H.M.; Hurley, W.L.
 Comp. Biochem. Physiol. B 93, 929-934, 1989
 A:Title: Purification and characterization of bovine lactoferrin from secretions of the
 A:Reference number: PLO148; MUID:90031466; PMID:2805645
 A:Accession: PLO148
 A:Molecule type: protein
 A:Residues: 20-27, 'X', 29-37, 'X', 39-54, 'X', 56-59 <REU>
 R:Beilamy, W.; Takase, M.; Yamauchi, K.; Wakabayashi, H.; Kawase, K.; Tomita, M.
 Biochim. Biophys. Acta 1121, 130-136, 1992
 A:Title: Identification of the bactericidal domain of lactoferrin.
 A:Reference number: S21756; MUID:92287941; PMID:1599934
 A:Accession: S21756
 A:Molecule type: protein
 A:Residues: 36-60 <BEL>
 R:Shimazaki, K.; Tanaka, T.; Kon, H.; Oota, K.; Kawaguchi, A.; Maki, Y.; Sato, T.
 J. Dairy Sci. 76, 946-955, 1993
 A:Title: Separation and characterization of the C-terminal half molecule of bovine lacto
 A:Reference number: A56659; MUID:93253156; PMID:8486845
 A:Accession: A56659
 A:Molecule type: protein
 A:Residues: 20-25; 302-308; 359-366, 'X', 368-376, 'X', 378 <SHI>
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-708/Product: lactoferrin repeat homology <TRH1>
 F:36-60/Region: antitumor
 F:359-696/Domain: transferrin repeat homology <TRH2>
 F:28-64, 134-217, 176-192, 179-200, 189-202, 250-264, 367-399, 377-390, 424-703, 444-666, 476-551,
 F:38-55/Diulfide bonds: #status predicted
 F:79, 111, 211, 272/Binding site: iron (Asp, Tyr, Tyr, His) #status experimental
 F:40/Binding site: carbonate (Arg) #status experimental
 F:252, 300, 387, 495, 564/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:414, 452, 545, 613/Binding site: iron (Asp, Tyr, Tyr, His) #status experimental
 F:482/Binding site: carbonate (Arg) #status experimental

Query Match 100.0%; Score 47; DB 1; Length 708;
 Best Local Similarity 100.0%; Pred. No. 0.25;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 APRKXVW 8
 |||||
 Db 20 APRKXVW 27

RESULT 3
 JJC323
 lactoferrin - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 07-May-1999
 C:Accession: JJC323
 R:Le Provost, F.; Nocard, M.; Guerlin, G.; Martin, P.
 Biochem. Biophys. Res. Commun. 203, 1324-1332, 1994
 A:Title: Characterization of the goat lactoferrin cDNA: assignment of the relevant locus
 A:Reference number: JJC323; MUID:94380047; PMID:8093048
 A:Accession: JJC323
 A:Molecule type: mRNA
 A:Residues: 1-708 <LEP>
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein
 F:359-696/Domain: transferrin repeat homology <TRH2>
 F:252, 300, 387, 495, 564/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 47; DB 2; Length 708;
 Best Local Similarity 100.0%; Pred. No. 0.25;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 APRKXVW 8
 |||||
 Db 20 APRKXVW 27

RESULT 4
 A4543
 lactoferrin precursor - pig
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 04-Mar-2000
 C:Accession: A4543; S24173
 R:Alexander, L.J.; Levine, W.B.; Teng, C.T.; Beattie, C.W.
 Anim. Genet. 23, 251-256, 1992
 A:Title: Cloning and sequencing of the porcine lactoferrin cDNA.
 A:Reference number: A4543; MUID:92367939; PMID:1503259
 A:Accession: A4543
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-703 <ALE>
 A:Experimental source: mammary gland
 A:Note: sequence extracted from NCBI backbone (NCBIN:111151, NCBI:111153)
 R:Iyodon, J.P.; O'Malley, B.R.; Saucedo, O.; Lee, T.; Headon, D.R.; Connely, O.M.
 Biochim. Biophys. Acta 1132, 97-99, 1992
 A:Title: Nucleotide and primary amino acid sequence of porcine lactoferrin.
 A:Reference number: S24173; MUID:92379101; PMID:1511016
 A:Accession: S24173
 A:Molecule type: mRNA
 A:Residues: 1-11, 'W', 13-50, 'I', 52-84, 'G', 86-120, 'L', 121-130, 'I', 132-282, 'S', 284-571, 'Q',
 A:Cross-references: EMBL:92089; NID:g164613; PION:AAA31102.1; PID:g164614
 A:Experimental source: mammary gland
 C:Superfamily: transferrin; transferrin repeat homology
 C:Keywords: duplication; glycoprotein; iron; iron binding; metalloprotein; milk
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-703/Product: lactoferrin repeat homology <TRH1>
 F:20-350/Domain: transferrin repeat homology <TRH2>
 F:36-48/Region: antitumor
 F:354-691/Domain: transferrin repeat homology <TRH2>
 F:28-62, 38-53, 129-212, 171-187, 184-195, 245-259, 362-394, 372-385, 419-698, 439-661, 471-546, 49
 F:77, 107, 206, 267/Binding site: iron (Asp, Tyr, Tyr, His) #status predicted
 F:103/Binding site: carbonate (Arg) #status predicted
 F:409, 447, 540, 609/Binding site: iron (Asp, Tyr, Tyr, His) #status predicted
 F:477/Binding site: carbonate (Arg) #status predicted
 F:490/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 80.9%; Score 38; DB 2; Length 703;
 Best Local Similarity 75.0%; Pred. No. 15;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 |||:|
 20 APRKNVRW 27

RESULT 5

hypothetical protein C32B5.16 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T25580

R:Sheet, P.

submitted to the EMBL Data Library, December 1996

A:Description: The sequence of C. elegans cosmid C32B5.

A:Reference number: Z20054

A:Accession: T25580

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-345 <SCH>

A:Cross-references: EMBL:U80843; PIDN:AA37971.1; GSPDB:GN00020; CESP:C32B5.16

A:Experimental source: strain Bristol N2; clone C32B5

C:Genetics:

A:Gene: CESP:C32B5.16

A:Map position: 2

A:introns: 24/2, 304/2

C:Superfamily: Caenorhabditis elegans hypothetical protein F4262.3

Query Match 74.5%; Score 35; DB 2; Length 345;
 Best Local Similarity 83.3%; Pred. No. 29;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RKNVRW 8
 |||:|
 100 RKNVRW 105

RESULT 6

phm protein [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

C:Accession: AD3390

R:Belvecchio, V.G.; Kapetral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, P.; Masur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3252; PMID:11756688

A:Accession: AD3390

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-379 <KUR>

A:Cross-references: GB:AE008917; PIDN:AAL52287.1; PID:gl7983077; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME1106

A:Map position: 1

Query Match 74.5%; Score 35; DB 2; Length 379;
 Best Local Similarity 75.0%; Pred. No. 32;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 |||:|
 66 APRKNVRW 73

RESULT 7

pyrimidine nucleoside transport protein nupc [imported] - Staphylococcus aureus (strain

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: B89819

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shida, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: B89819

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-404 <KUR>

A:Cross-references: GB:BA000018; PID:g13700411; PIDN:BA841709.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: nupc

C:Superfamily: pyrimidine nucleoside transport protein nupc

Query Match 74.5%; Score 35; DB 2; Length 404;
 Best Local Similarity 83.3%; Pred. No. 34;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RKNVRW 8
 |||:|
 24 RKNVRW 29

RESULT 8

nonstructural polypeptide - Sindbis virus

N:Contains: nonstructural NS2; nonstructural protein NS1; nonstructural protein NS3; non

C:Species: Sindbis virus

C:Date: 19-Feb-1984 #sequence_revision 03-Aug-1984 #text_change 04-Oct-1996

C:Accession: A03917

R:Strauss, E.G.; Rice, C.M.; Strauss, J.H.

Virology 133, 92-110, 1984

A:Title: Complete nucleotide sequence of the genomic RNA of Sindbis virus.

A:Reference number: A94331; MUID:84148439; PMID:6322438

A:Accession: A03917

A:Molecule type: genomic RNA

A:Residues: 1-2512 <STR>

A:Experimental source: strain HRSP

A:Note: readthrough of the terminator UGA between codons UAC for 1896-Tyr and CUA for 18

C:Superfamily: Semliki Forest virus nonstructural protein

C:Keywords: polypeptide

F:1-540/Product: nonstructural protein NS1 #status predicted <NS1>

F:541-1347/Product: nonstructural protein NS2 #status predicted <NS2>

F:1348-1896/Product: nonstructural protein NS3 #status predicted <NS3>

F:1897-2512/Product: nonstructural protein NS4 #status predicted <NS4>

Query Match 74.5%; Score 35; DB 1; Length 2512;
 Best Local Similarity 62.5%; Pred. No. 2e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 |||:|
 1195 APRKNVRW 1202

RESULT 9

minor core protein - human adenovirus 12

C:Species: Mastadenovirus h12 (human adenovirus 12)

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C:Accession: S33939

R:Sprenkel, J.

submitted to the EMBL Data Library, June 1993

A:Reference number: S33928

A:Accession: S33939

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-347 <SPR>

A/Cross-references: EMBL:X73487; NID:G313361; PIDN:CAA51888.1; PID:G313373
C/Superfamily: adenovirus minor core protein pV

Query Match 72.3%; Score 34; DB 2; Length 347;
Best Local Similarity 62.5%; Pred. No. 47;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
| | | | |
Db 61 APRRRVQW 68

RESULT 10

FOADM5

minor core protein pV - human adenovirus 5

C/Species: Mastadenovirus h5 (human adenovirus 5)

A/Note: host Homo sapiens (man)

C/Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999

C/Accession: C39449

R/Chroboczek, J.; Bieber, F.; Jacrot, B.

Virolology 186, 280-285, 1992

A/Title: The sequence of the genome of adenovirus type 5 and its comparison with the gen

A/Reference number: A39449; MUID:92087470; PMID:11727603

A/Accession: C39449

A/Molecule type: DNA

A/Residues: 1-368 <CHR>

A/Cross-references: GB:M73260; GB:M29978; NID:G209842; PIDN:AAA96409.1; PID:G209846

C/Superfamily: adenovirus minor core protein pV

C/Keywords: core protein; late protein

Query Match 72.3%; Score 34; DB 1; Length 368;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
| | | | |
Db 59 APRRRVQW 66

RESULT 11

FOADM2

minor core protein pV - human adenovirus 2

C/Species: Mastadenovirus h2 (human adenovirus 2)

A/Note: host Homo sapiens (man)

C/Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 04-Mar-1994

C/Accession: A03837

R/Allestrom, P.; Akusjærvi, G.; Lager, M.; Yeh-kai, L.; Pettersson, U.

J. Biol. Chem. 259, 13980-13985, 1984

A/Title: Genes encoding the core proteins of adenovirus type 2.

A/Reference number: A03837; MUID:85054835; PMID:6094534

A/Accession: A03837

A/Molecule type: DNA

A/Residues: 1-369 <ALB>

C/Superfamily: adenovirus minor core protein pV

C/Keywords: core protein; late protein

Query Match 72.3%; Score 34; DB 1; Length 369;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
| | | | |
Db 60 APRRRVQW 67

RESULT 12

T11882

hypothetical protein a381r - Chlorella virus PBCV-1

C/Species: Chlorella virus PBCV-1

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T11882

R/Graves, M.V.; Van Etten, J.L.

submitted to the EMBL Data Library, May 1999

A/Reference number: Z18806

A/Accession: T117882

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-85 <GRA>

A/Cross-references: EMBL:U42580; NID:G4028896; PIDN:AAC96749.1

A/Experimental source: specific host Chlorella strain NC64A

C/Genetics:

A/Note: a381r

Query Match 70.2%; Score 33; DB 2; Length 85;
Best Local Similarity 75.0%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
| | | | |
Db 66 APRKNFRY 73

RESULT 13

T30991

hypothetical protein C01G5.7 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 04-Mar-2000

C/Accession: T30991

R/Bradshaw, H.; Stelliys, L.

submitted to the EMBL Data Library, August 1999

A/Description: The sequence of C. elegans cosmid C01G5.

A/Reference number: Z20956

A/Accession: T30991

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-164 <BRA>

A/Cross-references: EMBL:U50068; PIDN:AAB37737.1

A/Experimental source: strain Bristol N2

C/Genetics:

A/Map position: IV

A/Introns: 26/1; 70/1; 125/1

A/Note: C01G5.7

C/Superfamily: Caenorhabditis elegans hypothetical protein C01G5.7

Query Match 70.2%; Score 33; DB 2; Length 164;
Best Local Similarity 83.3%; Pred. No. 36;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RKNVRW 8
: | | | | |
Db 155 RKNVRW 160

RESULT 14

RGHUT1A

regenerating islet lectin 1-alpha precursor [validated] - human

N/Alternate names: lithostathine; pancreatic thread protein (PTP); reg I protein; reg1-a

N/Contains: pancreatic stone protein (PSP)

C/Species: Homo sapiens (man)

C/Date: 31-Mar-1990 #sequence_revision 03-Aug-1995 #text_change 08-Dec-2000

C/Accession: A35197; B28351; S12950; S02767; S02419; S00113; S01471; A25246

R/Matanahe, T.; Yonekura, H.; Terazono, K.; Yamamoto, H.; Okamoto, H.

J. Biol. Chem. 265, 7432-7439, 1990

A/Title: Complete nucleotide sequence of human reg gene and its expression in normal and

product of the gene.

A/Reference number: A35197; MUID:90237042; PMID:2332435

A/Accession: A35197

A/Molecule type: DNA

A/Residues: 1-166 <MAT>

A/Cross-references: GB:J05412

R/Terazono, K.; Yamamoto, H.; Takasawa, S.; Shiga, K.; Yonemura, Y.; Tochino, Y.; Okamoto

J. Biol. Chem. 263, 2111-2114, 1988

A/Title: A novel gene activated in regenerating islets.

A/Reference number: A92704; MUID:88115343; PMID:2963000

A/Accession: B28351

A/Molecule type: mRNA

A:Residues: 1-166 <TER>
 A:Cross-references: GB:M1963; NID:G190978; PIDN:AAA6558.1; PID:G190979
 R:Itch, T.; Tsunuki, H.; Katoh, T.; Teraoka, H.; Matsumoto, K.; Yoshida, N.; Terazono, K
 FBS Lett. 272, 85-88, 1990
 A:Title: Isolation and characterization of human reg protein produced in Saccharomyces c
 A:Reference number: S12950; MUID:91032149; PMID:2226837
 A:Accession: S12950
 A:Molecule type: protein
 A:Residues: 23-52;160-166 <ITD>
 A:Note: sequence determined from protein isolated after human cDNA sequence was cloned a
 R:de Caro, A.M.; Adich, Z.; Fournet, B.; Capon, C.; Bonicel, J.J.; de Caro, J.D.; Rover
 Biochim. Biophys. Acta 994, 281-284, 1989
 A:Title: N-terminal sequence extension in the glycosylated forms of human pancreatic stc
 A:Reference number: S02767; MUID:89150292; PMID:2493268
 A:Accession: S02767
 A:Molecule type: protein
 A:Residues: 23-47 <DEC>
 R:Rouimi, P.; de Caro, J.; Bonicel, J.; Rovey, M.; de Caro, A.
 FBS Lett. 229, 171-174, 1988
 A:Title: The disulfide bridges of the immunoreactive forms of human pancreatic stone pro
 A:Reference number: S02419; MUID:88152214; PMID:3345835
 A:Accession: S02419
 A:Molecule type: protein
 A:Residues: 63-72;125-139;150-157;160-166 <ROU>
 A:Note: disulfide bonds
 R:de Caro, A.M.; Bonicel, J.J.; Rouimi, P.; de Caro, J.D.; Sarles, H.; Rovey, M.
 Eur. J. Biochem. 168, 201-207, 1987
 A:Title: Complete amino acid sequence of an immunoreactive form of human pancreatic ston
 A:Reference number: S00113; MUID:88029417; PMID:3665916
 A:Accession: S00113
 A:Molecule type: protein
 A:Residues: 34-166 <DEI>
 R:Rouimi, P.; Bonicel, J.; Rovey, M.; de Caro, A.
 FBS Lett. 216, 195-199, 1987
 A:Title: Cleavage of the Arg-116 bond in the native polypeptide chain of human pancreati
 A:Reference number: S01471; MUID:87219142; PMID:3108036
 A:Accession: S01471
 A:Molecule type: protein
 A:Residues: 33-48 <RO2>
 R:Montalvo, G.; Bonicel, J.; Mulligner, L.; Rovey, M.; Sarles, H.; De Caro, A.
 Biochem. J. 238, 227-232, 1986
 A:Title: Partial amino acid sequence of human pancreatic stone protein, a novel pancreat
 A:Reference number: A25246; MUID:87099950; PMID:3541906
 A:Accession: A25246
 A:Molecule type: protein
 A:Residues: 34-73; 'X', 75-87; 'R', 89-98 <MON>
 C:Comment: This protein is found in pancreatic calculi of mammals. The tryptic-like clea
 C:Comment: Intact regenerating islet lectin 1-alpha, lithostathine, inhibits the growth
 C:Genetics:
 A:Gene: GDB:REG1A; REG
 A:Cross-references: GDB:132455; OMIM:167770
 A:Map position: 2p12-2p12
 A:Introns: 22/1; 61/3; 107/3; 145/1
 C:Superfamily: tetraneurin; C-type lectin homology
 C:Keywords: glycoprotein; lectin; pancreas; pyroglytamic acid
 F:1-22/Dominant: signal sequence #status predicted <SIG>
 F:23-166/Product: regenerating islet lectin 1-alpha #status experimental <MAT>
 F:34-166/Product: pancreatic stone protein #status experimental <MAT2>
 F:36-162/Dominant: C-type lectin homology <LCH>
 F:23/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen
 F:27/Binding site: carboxylate (Thr) (covalent) #status experimental
 F:33-34/Cleavage site: Arg-116 (trypsin) #status experimental
 F:36-47,64-162,137-154/Disulfide bonds: #status experimental

Query Match 70.2%; Score 33; DB 1; Length 166;
 Best Local Similarity 71.4%; Pred. No. 36;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRKNVRW 8
 DB 105 PKKNRRW 111

RESULT 15
 REG1UB
 regenerating islet lectin 1-beta precursor - human
 N:Alternate names: reg-related protein; reg1-beta protein
 M:Comments: pancreatic stone protein (PSP)
 C:Species: Homo sapiens (man)
 C:Date: 22-Nov-1993 #sequence revision 03-Aug-1995 #text_change 16-Jun-2000
 C:Accession: S34591; S42729; A44712
 R:Baroli, C.; Charib, B.; Georgi, D.; Sansonetti, A.; Dagorn, J.C.; Borge-LeFranc, J.L
 FBS Lett. 327, 289-293, 1993
 A:Title: A gene homologous to the reg gene is expressed in the human pancreas.
 A:Reference number: S34591; MUID:93351647; PMID:8346956
 A:Accession: S34591
 A:Molecule type: DNA
 A:Residues: 1-166 <BAR>
 A:Cross-references: GB:L08010; NID:G307368; PIDN:AAA18204.1; PID:G487726
 A:Note: this gene appears to be expressed in pancreas and liver
 R:Morizumi, S.; Watanabe, T.; Umno, M.; Nakagawara, K.; Suzuki, Y.; Miyashita, H.; Yone
 Biochim. Biophys. Acta 1217, 199-202, 1994
 A:Title: Isolation, structural determination and expression of a novel reg gene, human r
 A:Reference number: S42729; MUID:94153997; PMID:8110835
 A:Accession: S42729
 A:Molecule type: mRNA
 A:Residues: 1-166 <MOB>
 A:Cross-references: GB:D16816; NID:G474305; PIDN:BA04091.1; PID:G474306
 A:Accession: A44712
 A:Molecule type: DNA
 A:Residues: 1-166 <MO2>
 A:Cross-references: GB:D17291; NID:G474307; PIDN:BA04124.1; PID:G474308
 C:Comment: This protein is found in pancreatic calculi of mammals. The tryptic-like clea
 C:Genetics:
 A:Gene: GDB:REG1B; REG1
 A:Cross-references: GDB:342079
 A:Map position: 2p12-2p12
 A:Introns: 22/1; 61/3; 107/3; 145/1
 C:Superfamily: tetraneurin; C-type lectin homology
 C:Keywords: glycoprotein; lectin; pancreas; pyroglytamic acid
 F:1-22/Dominant: signal sequence #status predicted <SIG>
 F:23-166/Product: regenerating islet lectin 1-beta #status predicted <MAT>
 F:34-166/Product: pancreatic stone protein #status predicted <MAT2>
 F:36-162/Dominant: C-type lectin homology <LCH>
 F:23/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F:27/Binding site: carboxylate (Thr) (covalent) #status predicted
 F:33-34/Cleavage site: Arg-116 (trypsin) #status predicted
 F:36-47,64-162,137-154/Disulfide bonds: #status predicted

Query Match 70.2%; Score 33; DB 1; Length 166;
 Best Local Similarity 71.4%; Pred. No. 36;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PRKNVRW 8
 DB 105 PKKNRRW 111

Search completed: July 30, 2003, 16:31:22
 Job time : 8.09524 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

Om protein - protein search, using sw model

Run on: July 30, 2003, 16:09:50 ; Search time 15.7619 Seconds
(without alignments)
80.562 Million cell updates/sec

Title: US-09-787-070-8
Perfect score: 47
Sequence: 1 APRKNVRW 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 11078633
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :
  A_GeneSeq_19Jun03.*
1: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/genseq/genseqp-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/genseq/genseqp-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/genseq/genseqp-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/genseq/genseqp-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/genseq/genseqp-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	47	100.0	8	21	AAV88343	N-terminal peptide
2	47	100.0	18	17	AAE91193	lactoferrin decomp
3	47	100.0	51	17	AAE91194	lactoferrin decomp
4	47	100.0	52	16	AAE87909	Bovine lactoferrin
5	47	100.0	689	21	AAV49370	lactoferrin sequen
6	47	100.0	704	22	AAW51535	Recombinant lactof
7	47	100.0	708	17	AAW09343	Bovine lactoferrin
8	47	100.0	708	19	AAW57318	Bovine lactoferrin
9	47	100.0	708	20	AAW86022	Bovine lactoferrin

10	47	100.0	708	21	AA808183
11	47	100.0	708	22	AA648827
12	47	100.0	708	22	AAE02344
13	38	80.9	685	12	AA811663
14	38	80.9	685	12	AA811663
15	38	80.9	703	17	AA809344
16	38	80.9	703	19	AA853139
17	38	80.9	703	20	AA866022
18	38	80.9	703	21	AA808183
19	38	80.9	703	22	AAE02344
20	37	78.7	137	22	AA800538
21	37	78.7	323	22	AA800957
22	37	78.7	380	23	ABB90775
23	37	78.7	380	24	AB854482
24	37	78.7	560	23	ABP51734
25	36	76.6	708	20	AA807774
26	36	76.6	708	20	AA807774
27	35	74.5	61	22	AA806759
28	35	74.5	404	22	AA884055
29	35	74.5	406	23	ABP32323
30	35	74.5	807	19	AA870466
31	34	72.3	98	23	ABP08722
32	34	72.3	99	23	AB809566
33	34	72.3	170	22	AA806052
34	34	72.3	305	21	AA856161
35	34	72.3	337	22	AA860661
36	34	72.3	356	24	ABP56644
37	34	72.3	389	22	AAU42093
38	34	72.3	439	21	AA849085
39	34	72.3	439	22	AA893327
40	34	72.3	439	22	AAU44447
41	34	72.3	439	23	ABU05543
42	34	72.3	439	24	AB847555
43	34	72.3	439	24	ABU70703
44	34	72.3	439	24	AB884223
45	34	72.3	19398	24	ABB96399

ALIGNMENTS

	RESULT 1
ID	AA7869343 standard; peptide; 8 AA
XX	AA786343
AC	AA786343;
XX	
DT	14-JUL-2000 (first entry)
XX	
DE	N-terminal peptide derived bovin
XX	
KW	Peptide production; biological f
KW	antibacterial peptide; lactoferr
XX	
OS	Bos sp.
XX	
PN	WO20015655-A1.
XX	
PD	23-MAR-2000.
XX	
PF	15-SEP-1999; 99WO-BP07002.
XX	
PR	15-SEP-1998; 98EP-0203107.
PR	08-JUN-1999; 99EP-0201815.
XX	
PA	(NIZO-) NIZO FOOD RES.
XX	
PI	Vissers S, Recio I;
XX	
DR	WP1; 2000-271377/23.
PT	
FT	Novel process for producing pept
	from biological fluids such as m

Amino acid sequenc
 Chronic hepatitis
 Bovine lactoferrin
 Partial porcine/hu
 Porcine lactoferrit
 Porcine lactoferrit
 Porcine lactoferrit
 Amino acid sequenc
 Porcine lactoferrit
 Human polypeptide
 Human bone marrow
 Human Tumour Endoc
 Human normal endoc
 Human MDDR SEQ ID
 Goat lactoferrin-AB
 Goat lactoferrin-AB
 Proponibacterietum
 S. epidermidetis ope
 Straphylococcus epi
 Sindbis virus ns2/2
 Human ORX protein
 Human apoptosis re
 Proponibacterietum
 Human prostatic can
 Human apoptosis-re
 Chimpanzee C66 adad
 Human novel protea
 A human proliferat
 Human polypeptide
 Human novel protib
 Breast cancer-asso
 Breast cancer-asso
 Human breast cancer
 Adenovirus Adip
 Streptomyces virid

PT fluid with chromatographic medium to adsorb peptide domain of interest
 PT
 XX
 PS Claim 14; Page 22; 41pp; English.

CC This sequence represents an N-terminal peptide derived from bovine whey.
 CC The peptide is an example of a peptide with antibacterial activity that
 CC can be produced by the process of the invention. The invention relates to
 CC a process for producing peptides from biological fluids. The process
 CC comprises chromatography of the biological fluid, in situ hydrolysis of
 CC selectively bound peptides, washing to remove unbound peptide, and
 CC elution of the peptides of interest. The process is used for producing
 CC peptides from biological fluids, such as milk, whey or blood. For
 CC example, the process can be used to produce antibacterial peptides
 CC derived from lactoferrin, using cheese whey as a starting material. The
 CC peptides obtained have preferably antimicrobial and/or antiviral and/or
 CC antitumor activity. The process of the invention is relatively simple
 CC and generally economically and technically more attractive than those
 CC methods previously used. The method provides high yield peptides with a
 CC selected activity of interest without the need for intermediate
 CC purification of the precursor protein.

XX
 SQ Sequence 8 AA;

Query Match 100.0%; Score 47; DB 21; Length 8;
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 |||||
 DB 1 APRKNVRW 8

RESULT 2
 ID AAR91193 standard; peptide; 18 AA.

XX AAR91193;

DT 06-SEP-1996 (first entry)

DE Lactoferrin decomposition peptide.

XX Immunoadjuvant; antiviral; cytomagalovirus; cosmetic; food; feed;

KM lymphocyte blastogenesis.

XX Synthetic.

PN JP08073499-A.

PD 19-MAR-1996.

PF 01-SEP-1994; 94JP-0232026.

PR 01-SEP-1994; 94JP-0232026.

PA (SNOW) SNOW BRAND MILK PROD CO LTD.

DR WPI; 1996-205535/21.

PT New peptide(s) derived from human lactoferrin - are useful as
 PT immuno:activators, esp. for preventing infection by cytomagalovirus

PS Claim 4; Page 2; 11pp; Japanese.

CC The present sequence is one of four new peptides (see AAR91191 -
 CC AAR91193) obtained by enzymatically decomposing lactoferrin using
 CC protease. The peptides are immunoactivators which induce
 CC blastogenesis of lymphocytes and are particularly useful for
 CC treating cytomagalovirus infection. They can be used in drugs and
 CC cosmetics and can be added to foods and feeds.

XX
 SQ Sequence 18 AA;

Query Match 100.0%; Score 47; DB 17; Length 18;
 Best Local Similarity 100.0%; Pred. No. 0.096;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 |||||
 DB 1 APRKNVRW 8

RESULT 3
 ID AAR91194 standard; peptide; 51 AA.

XX AAR91194;

DT 06-SEP-1996 (first entry)

DE Lactoferrin decomposition peptide.

XX Immunoadjuvant; antiviral; cytomagalovirus; cosmetic; food; feed;

KM lymphocyte blastogenesis.

XX Synthetic.

OS Key Location/Qualifiers

FT Disulfide-bond 9..45 /note= "optionally this bond may be reduced"

FT Disulfide-bond 19..36 /note= "optionally this bond may be reduced"

PN JP08073499-A.

PD 19-MAR-1996.

PF 01-SEP-1994; 94JP-0232026.

PR 01-SEP-1994; 94JP-0232026.

PA (SNOW) SNOW BRAND MILK PROD CO LTD.

DR WPI; 1996-205535/21.

PT New peptide(s) derived from human lactoferrin - are useful as
 PT immuno:activators, esp. for preventing infection by cytomagalovirus

PS Claim 5; Page 2; 11pp; Japanese.

CC The present sequence is one of four new peptides (see AAR91191 -
 CC AAR91193) obtained by enzymatically decomposing lactoferrin using
 CC protease. The peptides are immunoactivators which induce
 CC blastogenesis of lymphocytes and are particularly useful for
 CC treating cytomagalovirus infection. They can be used in drugs and
 CC cosmetics and can be added to foods and feeds.

XX
 SQ Sequence 51 AA;

Query Match 100.0%; Score 47; DB 17; Length 51;
 Best Local Similarity 100.0%; Pred. No. 0.23; 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 |||||
 DB 1 APRKNVRW 8

RESULT 4
 ID AAR87909 standard; peptide; 52 AA.

XX AAR87909;

DT 01-MAR-1996 (first entry)

XX	Bovine lactoferrin (1-52).
DE	
XX	antiviral; lactoferrin;
KW	
XX	Synthetic.
OS	
XX	
PH	Key
DH	Location/Qualifiers
FT	Disulfide-bond 9..45
FT	/note= "each Cys residue may be in reduced form or may
FT	form a disulphide bond with another Cys residue."
FT	A 9-45 disulphide bond is exemplified"
FT	Disulfide-bond 19..36
FT	/note= "each Cys residue may be in reduced form or may
FT	form a disulphide bond with another Cys residue."
FT	A 19-36 disulphide bond is exemplified"
PV	JP07069915-A.
PN	
XX	
PD	14-MAR-1995.
XX	
PF	02-SEP-1993; 93JP-0240284.
XX	
PR	02-SEP-1993; 93JP-0240284.
XX	
PA	(SNOW) SNOW BRAND MILK PROD CO LTD.
XX	
DR	WPI; 1995-144726/19.
XX	
PT	Inhibitor against viral infection and proliferation - contains
PT	peptide having sequence from lactoferrin
PS	
XX	Claim 5; Page 2; 10pp; Japanese.
CC	The sequence is one of six peptides disclosed as having inhibitory
CC	effect against viral infection. The peptides are derived from
CC	lactoferrin. Their activity is demonstrated against cytomegalovirus.
XX	
SQ	Sequence 52 AA;
Query Match	100.0%; Score 47; DB 16; Length 52;
Best Local Similarity	100.0%; Pred. No. 0.23;
Matches 8; Conservative	0; Mismatches 0; Indels 0; Gaps 0
Cy	1 APRKNVRW 8
Db	1 APRKNVRW 8
RESULT 5	
ID	AAAY49270
XX	AAAY49270 standard; protein; 689 AA.
AC	AAAY49270;
XX	
DT	07-FEB-2000 (first entry)
XX	
DE	Lactoferrin sequence.
KM	Adhesion inhibition; enteropathogenic; Escherichia coli; lactoferrin;
XX	medicament; drink; feed; food poisoning.
XX	
OS	Unidentified.
XX	
PN	JP11292789-A.
XX	
PD	26-OCT-1999.
XX	
PF	03-APR-1998; 98JP-0107167.
XX	
PR	03-APR-1998; 98JP-0107167.
XX	
PA	(MEIP) MEIJI MILK PROD CO LTD.

XX WP1; 2000-018674/02.

DR Adhesion inhibitory compositions of enteropathogenic E. coli to cells -

XX used in medicines and foodstuffs

PT Disclosure; Fig 4; 7pp; Japanese.

XX

XX The invention provides an adhesion inhibitory composition of

CC enteropathogenic Escherichia coli to cells. The composition contains a

CC lactoferrin degraded matter. The composition is useful as a medicament,

CC a drink and food or a feed. The composition is effective for prevention

CC and/or treatment of infectious food poisoning caused by enteropathogenic

CC E. coli.

XX

XX Sequence 689 AA;

SQ

Query Match 100.0%; Score 47; DB 21; Length 689;

Best Local Similarity 100.0%; Pred. No. 2,6; 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 APRKQVNW 8

Db 1 APRKQVNW 8

RESULT 6

AAM51535

ID AAM51535 standard; Protein; 704 AA.

XX

AC AAM51535;

XX

DT 07-JAN-2002 (first entry)

XX

DE Recombinant lactoferrin.

XX

KW Bovine; recombinant; lactoferrin; pWE; Candida methylca SKL60.

OS Bos sp.

XX

PN KR2001039000-A.

PD 15-MAY-2001.

XX

PF 28-OCT-1999; 99KR-0047203.

XX

PR 28-OCT-1999; 99KR-0047203.

XX

PA (EASY-) EASY BIO SYSTEM INC.

XX

PA (SUNG/) SUNG C G.

XX

PI Min DG, Mun TH, Sung CG;

XX

DR WP1; 2001-594660/67.

XX

DR N-PSDB; AAI69871.

XX

PT Manufacturing Candida methylca skL60 (Kctc 8968p) and recombinant

XX

XX lactoferrin -

XX

PS Disclosure; Fig 5; 15pp; Korean.

XX

XX The invention relates to a method for manufacturing a recombinant yeast

CC strain for lactoferrin expression. A recombinant vector pWE containing

CC the lactoferrin gene is used to transform Candida methylca SKL60,

CC which, in turn, produces recombinant lactoferrin at high yield.

CC The method of manufacturing a recombinant lactoferrin gene comprises:

CC (i) separating RNA from Korean cow, using antibacterial cDNA to make a

CC primer for amplifying antibacterial lactoferrin DNA, and inserting

CC the lactoferrin gene into vector pWE to make an expression vector pWE;

CC (ii) electrically shocking Candida methylca MXL-6 to insert the

CC expression vector pWE into the bacteria;

CC (iii) pre-inducing the transformed strain with 10 % of glycerol and

CC inducing the transformant with 0.5-6 % of ethanol; and

CC (iv) cultivating the bacteria in pH 4.0-6.0 of a fermenter at 30 plusoc
 CC for 14-20 hours.
 CC The present sequence is the lactoferrin polypeptide.

XX Sequence 704 AA;

Query Match 100.0%; Score 47; DB 22; Length 704;
 Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 DB 20 APRKNVRW 27

RESULT 7

ID AAM09343 standard; Protein; 708 AA.

XX AAM09343;

DT 25-MAR-2003 (updated)
 DT 18-MAR-1997 (first entry)

DE Bovine lactoferrin.

XX Human; lactoferrin; iron-binding glycoprotein; milk; secretion; fungus;
 KM transferrin; bactericidal activity; prostate; expression system; primer;
 KM PCR; polymerase chain reaction; amplification; signal peptide; antiviral;
 KM alpha-amylase; Aspergillus oryzae; nutrition; bovine.

OS Bos taurus.

XX US5571691-A.

PN 05-NOV-1996.

PD 28-OCT-1993; 93US-0145681.

XX 28-OCT-1993; 93US-0145681.

PR 05-MAY-1989; 89US-0348270.

PR 28-SEP-1989; 89US-0413880.

PR 24-APR-1992; 92US-0873304.

PR 27-OCT-1992; 92US-0967947.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

PI Connely OM, Headon DR, May GS, O'Malley BW;

DR MPI; 1996-505400/50.

DR N-PSDB; AAT48031.

XX New DNA encoding human lactoferrin, its natural alleles and

PT substitution analogues - useful e.g. for preventing iron deficiency

PT and as antiviral/antimicrobial agent

XX Disclosure; Column 31-34; 92pp; English.

PS This is the amino acid sequence of bovine lactoferrin, an iron-binding

CC glycoprotein found in milk and other secretions and body fluids.

CC Fragments of the protein are also known to have bactericidal activity e.g.

CC the N-terminal portion of the protein has a bactericidal activity. The

CC gene was used to construct a fusion protein in which the native

CC lactoferrin signal peptide was replaced by the alpha-amylase II signal

CC peptide. The novel construct was then expressed in Aspergillus oryzae.

CC The protein can be used for antibacterial and antiviral activities as

CC well as an iron-carrying protein for nutritional or therapeutic

CC applications.

CC (Updated on 25-MAR-2003 to correct PF field.)

CC Sequence 708 AA;

XX Query Match 100.0%; Score 47; DB 17; Length 708;

Best Local Similarity 100.0%; Pred. No. 2.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8
 DB 20 APRKNVRW 27

RESULT 8

ID AAM57318 standard; Protein; 708 AA.

XX AAM57318;

DT 10-AUG-1998 (first entry)

DE Bovine lactoferrin.

XX Bovine; lactoferrin; recombinant; therapeutic; nutritional; iron;

KM Fe binding site; bacteria; bactericidal; milk.

OS Bos taurus.

XX US5766939-A.

PN 16-JUN-1998.

PD 30-MAY-1995; 95US-0453703.

XX 28-OCT-1993; 93US-0145681.

PR 05-MAY-1989; 89US-0348270.

PR 24-APR-1992; 92US-0873304.

PR 27-OCT-1992; 92US-0967947.

PR 30-MAY-1995; 95US-0453703.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

PI Connely OM, Headon DR, May GS, O'Malley BW;

DR MPI; 1998-361705/31.

DR N-PSDB; AAV30771.

XX Plasmids containing human lactoferrin DNA - for recombinant

PT production of the enzyme, especially fragments having bactericidal

PT activity

XX Example 13; Fig 14; 92pp; English.

PS The present sequence represents bovine lactoferrin. A plasmid has been

CC developed which is suitable for the expression of a human lactoferrin,

CC or an iron-binding lobe of lactoferrin, in a transformed prokaryotic

CC host cell. The plasmid comprises a DNA sequence encoding a naturally

CC occurring human lactoferrin protein or an iron-binding lobe of

CC lactoferrin and further comprises transcriptional and translational

CC regulatory elements capable of regulating the expression of the

CC lactoferrin-encoding DNA sequence in the transformed host cell. The

CC plasmid is useful for producing recombinant human lactoferrin proteins

CC in bacteria. Lactoferrin is a 78 kDa iron-binding glycoprotein found

CC in milk and other secretory fluids. It is involved in iron transfer

CC and delivery in mammals. It has been implicated as a resistance

CC factor in suckled new born infants against enteric infections; the

CC iron binding activity thought to be responsible for bactericidal action

CC as the bound iron deprives microorganisms of the metal, and interferes

CC with their growth and reproduction. The protein can be used for

CC nutritional and therapeutic applications. It has not been previously

CC possible to efficiently produce lactoferrin recombinantly.

CC Sequence 708 AA;

XX Query Match 100.0%; Score 47; DB 19; Length 708;

XX Best Local Similarity 100.0%; Pred. No. 2.7;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	APRKNVRW	8
Db	20	APRKNVRW	27

Query Match	100.0%;	Score 47;	DB 21;	Length 708;
Best Local Similarity	100.0%;	Pred. No. 2.7;		
Matches	8;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0

QY 1 APRKNVRW 8
 DB 20 APRKNVRW 27

RESULT 11

AA64827 standard; protein; 708 AA.

AC AAG64827;

DT 20-SEP-2001 (first entry)

DE Chronic hepatitis treatment related protein SEQ ID NO: 8.

KM Chronic hepatitis; viral antigenic protein; hepatitis C; hepatitis B.

OS Bos taurus.

PN WO200147545-A1.

PD 05-JUL-2001.

PF 28-DEC-2000; 2000WO-JP09393.

PR 28-DEC-1999; 99JP-0374087.

PA (SUMO) SUMITOMO PHARM CO LTD.

PI Tohdoh N, Murata M, Enjoji T;

DR WPI; 2001-425585/45.

PT Treatment and prevention of chronic hepatitis

PS Example 1; Page 74-78; 128pp; Japanese.

CC The present invention describes a method of preventing and treating

CC chronic hepatitis, involving administering an oligopeptide which

CC (a) has binding affinity towards the viral antigenic protein;

CC (b) inhibits binding affinity of the virus towards the receptor protein

CC of the target cell and

CC (c) has analogy with the receptor protein at the amino acid level. This

CC is a protein described in the exemplification of the invention.

SO Sequence 708 AA;

Query Match 100.0%; Score 47; DB 22; Length 708;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8

DB 20 APRKNVRW 27

RESULT 12

AA602342 standard; protein; 708 AA.

AC AAE02342;

DT 10-AUG-2001 (first entry)

DE Bovine lactoferrin (LF).

KM Bovine; lactoferrin; LF; therapeutic; nutritional; iron transport;

KM virucidal; bactericidal; animal food; iron-binding glycoprotein.

OS Bos taurus.

PN US6228614-B1.

XX 08-MAY-2001.
 PD 09-MAR-1999; 99US-0265577.
 PF

XX 28-OCT-1993; 93US-0145681.
 PR 05-MAY-1989; 89US-0348270.
 PR 30-MAY-1995; 95US-0456108.
 PR 24-APR-1992; 92US-0873304.
 PR 27-OCT-1992; 92US-0967947.

PA (BAY) BAYLOR COLLEGE MEDICINE.

PI Conneely OM, Heaton DR, O'malley BW, May GS;

DR WPI; 2001-342673/36.

DR N-Psdb; AAD06282.

PT Producing recombinant human lactoferrin useful for therapeutic or

PT nutritional applications, comprises transforming a eukaryotic cells

PT with a vector having a DNA that encodes and permits expression of

PT lactoferrin in the cells

PS Example 13; Fig 14B; 92pp; English.

CC The present invention relates to a method for expressing human

CC lactoferrin in an eukaryotic cell. The method comprises introducing

CC a plasmid comprising DNA encoding a human lactoferrin protein and

CC regulatory elements necessary for the expression of DNA in the cell.

CC The method is used for preparing recombinant lactoferrin, making

CC available a source of protein for therapeutic and nutritional

CC applications, e.g., as therapeutic additives to enhance iron transport

CC and delivery and for virucidal or bactericidal qualities or as human

CC or animal food. The present method is an efficient and economical way

CC to produce human lactoferrin.

CC The present sequence is bovine lactoferrin (LF). LF is an

CC iron-binding glycoprotein found in milk, other secretions and body

CC fluids.

SO Sequence 708 AA;

Query Match 100.0%; Score 47; DB 22; Length 708;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APRKNVRW 8

DB 20 APRKNVRW 27

RESULT 13

AA11663 standard; protein; 685 AA.

AC AAR11663;

DT 04-JUL-1991 (first entry)

DE Partial porcine lactoferrin protein.

KM Porcine lactoferrin; iron transport; iron binding.

OS Sus scrofa.

PN Key Location/Qualifiers

FT Peptide 1..18 /label= porcine lactoferrin signal peptide

PN WO9105045-A.

PD 18-APR-1991.

PF 14-SEP-1990; 90WO-US05245.

XX 28-SEP-1989; 89US-0413880.
 PR (GRAN-) GRANADA BIOSCIENCES.
 PA
 XX
 PI Headon DR, Conneely OM, O'Malley BW;
 DR WPI; 1991-132852/18.
 DR N-PSDB; AAQ11559.
 XX
 PT New cDNA and amino acid sequence of porcine lactoferrin - has
 PT therapeutic and nutritional applications, especially iron
 PT transport and delivery
 PS
 XX Disclosure; fig 1; 19pp; English.
 CC This sequence comprises a portion of the porcine lactoferrin protein
 CC (most of signal peptide and part of mature protein incl. two iron-
 CC binding domains). This protein, or a constituent polypeptide, contg.
 CC at least one of the iron binding domains, is useful for eg improving
 CC an animals resistance to bacterial and viral infection. It can be used
 CC in a variety of prods. incl. feed, therapeutic additives and additives
 CC for eye drops, contact lens solns., etc. The partial porcine lacto-
 CC ferrin signal peptide can be replaced by a human lactoferrin signal
 CC peptide (see AAR11664).
 CC
 SQ Sequence 685 AA;

Query Match 80.9%; Score 38; DB 12; Length 685;
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 APRKNVVR 8
 ||:|||||
 DB 20 APKKGVRW 27

RESULT 14
 AAR11664
 ID AAR11664 standard; Protein; 685 AA.

AC AAR11664;
 DT 04-JUL-1991 (first entry)

DE Partial porcine/human lactoferrin fusion protein.

KW Porcine lactoferrin; iron transport; iron binding.

OS Sus scrofa.

XX Key Location/Qualifiers

FT Peptide 1..19 /label= human lactoferrin signal peptide

FT Protein 20..680 /label= partial mature porcine lactoferrin

PN MO9105045-A.

PD 18-APR-1991.

PF 14-SEP-1990; 90WO-US05245.

PR 28-SEP-1989; 89US-0413880.

PA (GRAN-) GRANADA BIOSCIENCES.

PI Headon DR, Conneely OM, O'Malley BW;

DR WPI; 1991-132852/18.

DR N-PSDB; AAQ11560.

PT New cDNA and amino acid sequence of porcine lactoferrin - has

PT therapeutic and nutritional applications, especially iron
 PT transport and delivery

PS Disclosure; fig 2; 19pp; English.

CC This sequence comprises a portion of the porcine lactoferrin prot-
 CC ein (incl. two iron binding domains). This protein, or a consti-
 CC uent polypeptide, contg. at least one of the iron binding domains,
 CC is useful for eg improving an animals resistance to bacterial and
 CC viral infection. It can be used in a variety of prods. incl. feed,
 CC therapeutic additives and additives for eye drops, contact lens
 CC solns., etc.
 CC See also AAR11663.
 CC
 SQ Sequence 685 AA;

Query Match 80.9%; Score 38; DB 12; Length 685;
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 APRKNVVR 8
 ||:|||||
 DB 20 APKKGVRW 27

RESULT 15
 AAM09344
 ID AAM09344 standard; Protein; 703 AA.

AC AAM09344;

DT 25-MAR-2003 (updated)
 DT 18-MAR-1997 (first entry)

DE Porcine lactoferrin.

KW Human; lactoferrin; iron-binding glycoprotein; milk; secretion; fungus;

KW transferrin; bactericidal activity; prostate; expression system; primer;

KW PCR; polymerase chain reaction; amplification; signal peptide; antiviral;

OS Sus scrofa.

PN US5571691-A.

PD 05-NOV-1996.

PF 28-OCT-1993; 93US-0145681.

PR 28-OCT-1993; 93US-0145681.

PR 05-MAY-1989; 89US-0348270.

PR 28-SEP-1989; 89US-0413880.

PR 24-APR-1992; 92US-0873304.

PR 27-OCT-1992; 92US-0967947.

PA (BAYLOR COLLEGE MEDICINE.

PI Conneely OM, Headon DR, May GS, O'Malley BW;

DR WPI; 1996-505400/50.

DR N-PSDB; AAT48032.

XX New DNA encoding human lactoferrin, its natural alleles and

PT substitution analogues - useful e.g. for preventing iron deficiency

PT and as antiviral/antimicrobial agent

PS Disclosure; Column 37-40; 92pp; English.

This is the amino acid sequence of porcine lactoferrin, an iron-binding glycoprotein found in milk and other secretions and body fluids. Fragments of the protein are also known to have biological activity e.g. the N-terminal portion of the protein has a bactericidal activity. The gene was used to construct a fusion protein in which the native

CC lactoferrin signal peptide was replaced by the alpha-amylase II signal
CC peptide. The novel construct was then expressed in *Aspergillus oryzae*.
CC The protein can be used for antibacterial and antiviral activities as
CC well as an iron-carrying protein for nutritional or therapeutic
CC applications.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX

SQ Sequence 703 AA;

Query Match 80.9%; Score 38; DB 17; Length 703;

Best Local Similarity 75.0%; Pred. No. 1.1e+02; Mismatches 1; Indels 0; Gaps 0;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 APRKNVW 8

Db 20 APRKGVW 27

Search completed: July 30, 2003, 16:23:31
Job time : 16.7619 secs